STATISTICS:
THEORY AND PRACTICE

M. K. GHOSH, M.A., B.COM., (Lond.), M. INST.
UNIVERSITY PROFESSOR AND HEAD OF THE DEPARTMENT OF
COMMERCE UNIVERSITY OF ALLAHABAD.
AUTHOR: TRANSPORT DEVELOPMENT AND CO-ORDINATION
JOINT AUTHOR: INSURANCE PRINCIPLES,
PRACTICE AND LEGISLATION

AND

S. C. CHAUDHRI, M.A., B.COM.
LECTURER IN THE DEPARTMENT OF COMMERCE, UNIVERSITY
OF ALLAHABAD. ALLAHABAD JUBILEE GOLD
MEDALIST AND EX-RESEARCH SCHOLAR,
ALLAHABAD UNIVERSITY

FOURTH EDITION RE-PRINT

(Revised and Enlarged)

ALLAHABAD
THE INDIAN PRESS, LTD.
1949

Printed and published by K. Mittra
at The Indian Press Ltd Allahabad

PREFACE TO THE FIRST EDITION

Statistics was once known as the Science of Kings, but now it has gained ground in almost every branch of human knowledge. For, the superstructure of human activity rests ultimately, if not primarily, upon a foundation of quantitative facts—facts, whose inherent complexity and confusion can be simplified and analyzed and which can be interpreted only with a knowledge of statistical methods. In this Age of Statistics, therefore, the importance of the study of the Science of Statistics cannot be over-emphasized, particularly for India whose development, in many spheres, is yet in its infancy. Indeed, the importance is being recognized, and the Indian universities have taken the lead in the matter. Naturally, the necessity of a suitable text-book on the subject for Indian students is more than made out.

This book is an attempt to furnish a simple, but comprehensive, text for those who desire to equip themselves with a knowledge of the elementary statistical methods to enable themselves to handle statistical problems like skilled workmen. It is, of course, primarily intended for the benefit of those interested in Economics, Commerce, Sociology or Administration, but the general principles it comprises of will be suited equally well to every other variety of statistical data.

In a book such as this, the use of the viewpoints and materials of other works of the parallel and higher standards is unavoidable. And, indeed, such works have been our valuable guide. But we have made every effort to so synthesize all these materials as to bring about unity and harmony. The treatment is non-mathematical, chiefly because a majority of those for whom this book is primarily meant are not expert mathematicians, and also because

we feel there is a necessity of fundamental exposition of the non mathematical, nonetheless vital, processes involved in statistical inquiries, analysis and interpretation. Naturally therefore, a fuller discussion of topics like Probability, Sampling, Regression, etc., which require mathematical treatment, could not be included. Once the readers have overcome their feeling of un familiarity and grasped the basic principles, it will be easy for them to pick up the higher and more mathematical statistics The discussion on Statistical Material in India and on Indian Index Numbers does not pretend to be exhaustive, but is designed to make the Indian student look around him. Special care has been taken to select exercises for each chapter suited to the M Com., M.A. and B Com, standards of different universities of India.

Our thanks are due to Mr Sham Bahadur Kodasi, M A B Com., who helped us in correcting the proofs. We shall be thankful for any suggestion to increase the usefulness of the book

DEPARTMENT OF COMMERCE,
UNIVERSITY OF ALLAHABAD.
November, 1943.

MOHIT KUMAR GHOSH
SUSHEEL CHANDRA CHAUDHRI

PREFACE TO THE SECOND EDITION

It has been an unexpected gratification to find that the response from several of the Universities and Colleges in India and other quarters to the first edition of this work was so hearty as to require a second edition of it in less than twelve months from publication. The fulfilment of this requirement has been delayed, among other reasons, by the scarcity of paper—so characteristic of the War-time conditions.

Opportunity has been taken of this new edition to make some additions, recast certain paragraphs and correct such errors as were noted in the first edition. The main additions are Indian Industrial Statistics, India's National Income, Method of Least-squares, Seasonal Variation Index, and Lag. The minor changes resulting from recasting the matter are too numerous to be detailed and run almost throughout the book. The intention in making them has been to treat the more difficult parts of the subject in greater detail than was done in the former work. The variety of illustrations and examples has been increased to render the technical portions capable of being more soundly grasped. A notable feature of this edition is the inclusion of a number of questions from the latest examination papers not only of the universities which were represented in the first edition but also of several others. In short, the changes incorporated are the most necessary part of those that the experience of us in teaching the subject to the under-graduate and post-graduate students at the University dictated, and it is hoped that the utility of these changes will be judged and appreciated by the reader as he goes through the work. It must be added that every care has been taken to retain the same design and purpose of this edition as were claimed for its predecessor.

Reviews by journals and opinions from teachers of the subject and others interested in it have been very kind and, in some cases equally suggestive. We are glad to avail ourselves of this opportunity of thanking them for their encouragement and look forward to its continuance

We crave the indulgence of the scrupulous reader for an error that may have crept in despite the laborious proof reading done by us

COMMERCE DEPARTMENT
ALLAHABAD UNIVERSITY
May, 1945

M. K. GHOSH
S. C. CHAUDHRI

CONTENTS

CHAPTER I

CHAPTER V

COLLECTION OF STATISTICAL DATA

CHAPTER VI

EDITING THE COLLECTED DATA

CHAPTER VII

STATISTICAL MATERIAL IN INDIA

CHAPTER VIII

CLASSIFICATION AND TABULATION OF DATA

CHAPTER IX

SIMPLE DERIVATIVES

CHAPTER X

STATISTICAL AVERAGES

CHAPTER XI

DISPERSION AND SKEWNESS

CHAPTER XII

INDEX NUMBERS

CHAPTER XIII

INDIAN AND FOREIGN INDEX NUMBERS

CHAPTER XIV

DIAGRAMMATIC REPRESENTATION

CHAPTER XV

GRAPHICAL PRESENTATION

CHAPTER XVI

ANALYSIS OF TIME SERIES

CHAPTER XX

INTERPRETATION OF DATA

STATISTICS: THEORY AND PRACTICE

CHAPTER I

GROWTH OF THE SCIENCE OF STATISTICS

The word 'Statistics' seems to have been derived from the Latin *Status*, or, the Italian *Statista*, both meaning a political state. In fact, the study of statistics had its origin in the compilation of facts and figures for purposes of administration of state. In this sense, the subject must have been in existence from very early times. In the days of yore the ruling chiefs used to take, as often as necessary, censuses of population and property within their domain to determine their man-power and material strength, and thereby planned their fiscal and military policies. Collection of data for other purposes, however, was not ruled out. Perhaps one of the earliest enumerations made was regarding the population and riches of Egypt, taken about 3050 B.C., to plan the erection of Pyramids. But the most common compilations during the Middle Ages were concerned with taxation, distribution of land and available soldiers. In India, administrative statistics were highly organized nearly two thousand years ago. Inscriptions and technical treatises abound in references to various kinds of statistics for the classic period of Sanskrit culture. A system of registration of births and deaths was enforced in Maurya India, while *Aîn-i-Akbari*—a great administrative and statistical survey of India—was compiled during the reign of Emperor Akbar. Past history of other countries also bears witness to the fact that statistics was originally concerned with matters of state and was regarded as the **Science of Statecraft.**

Mercantilistic Period

During the Mercantilistic period the policies of the Western European governments were directed to the dual purpose of encouraging such industries as enhanced the power of the state, and of securing a favourable balance of trade. This necessitated legislation for social, economic and political reforms which to be effective and adequate called for more comprehensive statistics than were considered sufficient during the Middle Ages. The bulk of statistical compilations consequently increased.

16th Century

The ancient astronomers contributed much to the propagation of the study of statistics. They compiled records of the motions of heavenly bodies and predicted about eclipses and positions of stars. Upon a study of the data collected by Tycho Brahe (1546-1601) Johannes Kepler discovered the three laws relating to the motion of planets on which the theory of gravitation was founded by Sir Isaac Newton. When the utility of statistical method for attaining the knowledge of nature was demonstrated enthusiasts in political, social and economic fields began resorting to a similar approach. Naturally therefore more and more data came to be collected.

17th Century

The seventeenth century opened with a new use for some of the compiled figures, viz. a study of vital and social statistics. In 1612 Professor George Obrecht, of Strasburg University illustrated how vital and criminal statistics could be utilized for devising plans to provide a system of life insurance and pensions and to reform the criminals. Captain John Graunt of London (1620-1674) made an analytical study in the realm of vital statistics in 1661. Casper Neumann studied the death records of Breslau in 1691 and prepared his notes and conclusions which fell into

the hands of Edmund Halley, the famous astronomer and scientist, through the Royal Society of London. Halley computed from them a complete life table, deduced the expectation of life at each age and paved the way for a scientific system of life insurance. Sir William Petty (1623-1687); also drew up and discussed mortality tables. Indeed, the first life insurance institution was founded in London in 1698.

18th Century—Statistics and Mathematics.

As the statistical data grew in abundance and many new fields were opened up for investigation, need was soon felt for improving upon the then crude and cumbersome methods of analysing and interpreting the figures. The labours of Petty and Halley had prepared the ground for a more scientific treatment of the statistical methods in the eighteenth century, which they received particularly at the hands of J. P. Süssmilch (1707-1767), a Prussian clergyman, who tried to demonstrate the doctrine of 'Natural Order' statistically in an important publication. Others devised statistical tables and geometric figures for purposes of comparison of data. But modern theory of statistics was, thus far, conspicuous by its absence. John Graunt, Petty and Süssmilch conducted their studies during the seventeenth and eighteenth centuries under the name of **political arithmetic** which functioned as eyes and ears of central government.

In the eighteenth century, however, an alliance was effected between statistics and mathematics, and foundations of the theory of probability were laid, when J. Bernoulli (1654-1705), a professor of Basel, mathematically elucidated the 'law of large numbers' in his work *Ars Conjectandi,* published posthumously, and Daniel Bernoulli (1700-1782) suggested the theory of 'moral expectation'. The subject of probability, it may be interesting to note, grew out of an analysis of hazards of those who played games of chance. Laplace, the noted scientist, who followed up

physical characteristics from one generation to another. His enquiry helped his great successor, Karl Pearson (1857-1936), to produce his notable work on biometry and to emphasize the indispensability of statistics for the evolutionist, as in his opinion the whole problem of evolution was a problem of statistics.

Statistics appeared rather late in the field of the science of economics, though a beginning was made by Sir William Petty in his work, *Political Arithmetic,* published in 1690, as also by Gregory King, who, about the same time, attempted to statistically demonstrate a relationship between supply of commodities and prices. By the eighteenth century valuable statistical material relating to population, occupations, taxes, agriculture, industry, trade, shipping etc., had been collected in most civilized countries; but there was no liasion whatsoever between statistical information and economic theory. Political economy was brought up in the Classical School, founded by Adam Smith, through his great work *Wealth of Nations,* published in 1776. Classical economists were staunch believers in the deductive and abstract method of reasoning, their lip-sympathy such as that held by J. S. Mill (1806-1873) to the advantages of statistical verification of deductive laws notwithstanding. W. S. Jevons (1835-1882), in his *Theory of Political Economy* published in 1871, also advocated verification of the deductive science of economics by the inductive science of statistics, and opined that political economy could be developed into an exact science if only commercial statistics were more complete and precise. Although Cournot (1801-1877), a renowned mathematical economist and writer on probability, did statistics a signal service by hinting at the application of the calculus of variations and making a first casual suggestion regarding the distinction between secular trend and periodic fluctuations, yet, it was W. S. Jevons who segregated seasonal movements, secular trends, and

tion of statistics is "first, to suggest empirical laws, which may or may not be capable of subsequent deductive explanation; and secondly, to supplement deductive reasoning by checking its results and submitting them to the test of experience."[2] It is now widely held that induction without deduction is barren, deduction without induction sterile.

Since the last decade of the last century two important factors have brought about a fundamental change in the place of statistics in economics. Since eighteen-eightees pure theory of statistics has made a remarkable improvement. Eminent men like August Meitzen, Francis Edgeworth, Francis Galton, Karl Pearson, G. Uduy Yule, C. B. Davenport, W. S. Gossett, A. L. Bowley, Adams, W. Pearsons, W. I. King and R, A. Fisher have done a great deal in advancing the theory far beyond its former limits. The development of statistical methods—of probability, sampling, curve-fitting, correlation, periodicity, and index-numbers—closely coincided with the enlargement of figurative data made possible by the establishment of statistical bureaus and scientific recording of population censuses in different countries of the world. These improvements in statistical material about the close of the nineteenth century mark the real inception of statistics in economics.

Thus there has grown up a kin-ship among mathematics, economics and statistics. The modern science of statistics is no longer synonymous with 'political arithmetic'. It has extended its scope to varied departments of human knowledge. It is concerned not merely with matters of state but also with the physical, biological, anthropological, meteorological, social, economic and other phenomena. Its methods are applied wherever a study of large numbers is involved.

---

[2] 2nd edition, page 338.

EXERCISES

(1) Trace the growth of the Science of Statistics and throw light on its future

(2) Mathematics has played in the past, as it does even today, a great part in Statistical Theory, and there could be no theory without it, but that theory is no more a branch of mathematics than is say engineering or astronomy—Discuss

(3) Explain the relationship between Economics and Statistics How far has the use of statistical methods in Economics led to its development?

(M A, Agra, 1942)

(4) Statistics was born in the needs of state administration, but is no longer concerned with matters of state—In the light of this statement, show how and what transformation has taken place in the meaning of Statistics

(5) Show the relationship between Statistics and Mathematics, and Statistics and Natural Sciences

(6) 'Statistics are the straw out of which I, like every other economist, have to make bricks.' (*Marshall*)

Explain in the light of the above observation, the relation between Economics and Statistics, and discuss how far it is correct to say that the Science of Economics is becoming statistical in its method

(M Com, Alld, 1944)

(7) 'When you can measure what you are speaking about and express it in numbers, you know something about it but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.'

(*Lord Kelvin*)

To what extent does Lord Kelvin's observation apply to economics?

(M.A., Agra, 1944).

CHAPTER II

DEFINITION OF STATISTICS

Statistics* are numerical statements of facts in any department of inquiry, placed in relation to each other[1]. Isolated unconnected figures are not statistics. 20, 38, 67, or 15, 32, 55 are undoubtedly quantitative figures but not statistics. For, neither do they concern a sphere of inquiry nor are they placed in relation to each other. But when we are told that for husbands in a certain community at ages 20, 38, 67 years and so on, the corresponding ages of wives are 15, 32, 55 years and so on, these figures at once become statistics. For now they throw light on the relationship between the ages of husbands and wives in the given community.

Characteristics of Statistics

1 Several facts emerge from the above.

Firstly, **statistics must be quantitatively expressed.** Qualitative expressions like 'young', 'middle-aged' and 'old' have been indicated by numerical expressions like 20, 38 and 67 or 15, 32 and 55 in the above example. Crops over a series of years expressed in maunds per acre are statistics, but expressed by such terms as 'good', 'fair', 'normal' or 'poor' are not.

Secondly, **statistics are always aggregates.** A single age of 20 or 38 years is not statistics, a series of ages are. A single birth, a sale or a consignment does not form statistics. Yet a number of births, sales, consignments are statistics, since they can be studied *in relation to time, place and frequency of occurrence.*

---

* The term statistics is applied to the science of statistics as well as to its subject matter. In the former sense it is used as singular, in the latter as plural, noun.

[1] Bowley, A. L., *An Elementary Manual of Statistics*, page 1.

Thirdly, **statistics should relate to a department of inquiry.** That is, the sphere on which they are to throw light must be definite and clear. Their purpose and object must be pre-determined. The purpose of a series of ages of husbands and wives in our example may be to find whether young husbands have young wives and the old, old.

Fourthly, **statistics must be capable of being placed in relation to each other.** That is, they should be comparable. To be so they should be homogeneous; they cannot be stray numerical facts, unrelated data, culled from indiscriminate sources having no common basis for selection. Ages of husbands are to be compared only with the corresponding ages of wives in our example.

But these are not the only characteristics which statistics possess. It should be added that **statistics are affected to a marked extent by multitude of causes.** They are hardly ever traceable to a single cause. They are related to other facts. The stature of a man, *e.g.*, is casually connected with his race, ancestry, diet, age, occupation, climate and habit.

It may be further added that **reasonable standards of accuracy must be observed in the enumeration or estimation of statistics,** if logical conclusions are to be based on them. Of course, the reasonableness of the standards will be determined by the purpose which statistics are to serve. If precise results are desired, statistics must be accurately compiled; but if only general impressions are looked for, even appreciable error may not matter.

A definition, comprehensive enough to include the above characteristics of statistical data, is given by Secrist,[2] who uses the term statistics as meaning **"aggregates of facts, 'affected to a marked extent by a multiplicity of causes,' numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a pre-determined purpose, and placed in relation to each other."**

---

[2] Secrist, H., *An Introduction to Statistical Methods*, 1933 ed., page 10.

a magnitude that it will swell up the general index of Indian business activity very high

Even the statistics of industrial production are not quite sufficient in India In spite of this *Capital* the well known weekly journal of Calcutta has been publishing every month an Index of Indian Industrial Activity since March 1938

Capital' Index of Indian Industrial Activity—This index is published monthly and 1935 is taken as the base year The series selected and the weights assigned to each item for computing this index are —

Series Selected	*Weight*
Industrial Production—	
1 Cotton Manufactures	9
2 Jute Manufactures	6
3 Steel Ingots	5
4 Pig Iron	8
5 Cement	5
6 Paper	3
Mineral Production—Coal	7
Rail & River borne Trade	24
Financial Statistics—Cheque Clearances	20
Trade Foreign & Coastal—	
Exports	4
Imports	3
Shipping Foreign & Coastal—	
Tonnage entered	3
Tonnage cleared	3

Since March 1941 Trade Foreign and Coastal and Shipping Foreign and Coastal have been left out Instead Notes in Circulation (base April 1935 to March 1936) with weight 6 and Consumption of Electricity with weight 7 have been included The weighted geometric mean forms the general index and seasonal fluctuations are eliminated by means of a twelve months moving average Index for

cement appeared up to 1937-38 and has since been discontinued with the remark 'figures not available'. A specimen of the construction of this index is given in table 17, Chapter X, page 161.

Statistics for the above series are taken from the monthly publications of the Department of Commercial Intelligence and Statistics and from Statistical Summary of the Reserve Bank of India. This index does not afford an idea of the activities of people living in rural areas. And, even so far as urban people are concerned it is not fully representative. It does not include the production of sugar, tea, hides and skins which are quite important in the Indian industrial structure to-day. However, in the absence of complete and adequate statistical data no better index could be compiled.[6]

BRITISH INDEX NUMBERS

Wholesale Price Index Numbers.

Three important wholesale price index numbers that are compiled and maintained in Great Britain are:—

(1) Board of Trade Index Number.
(2) *Economist* Index Number.
(3) *Statist* Index Number.

Board of Trade Index Number.—The present series relating to this index begins with January 1935 and replaces an older series dating from 1920, which had replaced a still older series designed before the last Great War. The total number of

[6] In addition to the index numbers discussed above indices of Physical Volume of Production, of Prices of Securities, and of Values of India's Exports and Imports are published in the *Monthly Survey of Business Conditions in India*. A new Food Index, with the week ending August 25, 1940, as the base has also been started since February, 1944. The *Capital* also compiles and publishes a monthly Stocks and Shares index.

statistics as "Classified facts respecting the condition of the people in a state especially those facts which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement".[3] This definition however, relates to the use of the term statistics in the sense of statistical data rather than in that of the science and even so is much in keeping with the original meaning of the term. In its modern sense the term is not confined to "the condition of the people in a state" but has stretched itself to almost every phenomenon—biological, astronomical, social, meteorological—where a study of large numbers is involved. Webster's definition is therefore inadequate.

According to Bowley, "Statistics is the science of the measurement of social organism regarded as a whole, in all its manifestations".[4] This definition, according to its author, concerns the student of sociology, political economy or demography. But when the author recognizes that "Statistics is not merely a branch of political economy nor is it confined to any one science",[5] its definition should not have been so drawn up as to limit its operations to only one field—*viz.* that of man and his activities. This definition is, therefore, not sufficiently inclusive.

Bowley further observes "Statistics may rightly be called the science of averages".[6] No doubt averages present a bird's-eye view of a mass of unintelligible data but there are other equally serviceable devices such as graphs, pictograms, correlation tables and co-efficients which modern statistics utilizes to comprehend the significance of the complex quantitative data. Therefore while this definition does not confine the scope of the science to a particular phenomenon, it is still inadequate in so far as it stresses only one of the several statistical methods.

---

[3] Quoted by King, W. I., *Elements of Statistical Methods*, page 20.
[4] Bowley, A. L., *Elements of Statistics*, page 7.
[5] *Ibid.*, page 3.
[6] *Ibid.*, page 7.

Suggesting a possible definition Bowley says that statistics may be called "the science of counting."[7] Analysing this definition he observes that while dealing with large numbers, such as a population census, counting is neither easy nor within the reach of an individual. Great numbers, instead of being counted, are estimated. Even estimation requires the co-operation of a group of people, since the numbers with which statistics concerns are very large. But because of varying degrees of intelligence and sense of accuracy among a group of workers, and also because of the difficulty of so clearly defining the object to be counted that every worker shall understand the same thing by the same definition, the estimates are not mathematically exact. Bowley then concludes that 'though all estimates of this nature are sometimes included under the term *Statistics,* this definition at once is too wide, and also does not bring out the distinctive nature of statistical method.'[8] Obviously, this definition suffers from the dual defect of emphasizing the method of counting used in arithmetic rather than that of estimation on which statistics so much relies, and of taking into account only the collection of data, leaving the analysis of the collected data quite out of consideration. Therefore, this definition is also far too restricted, though it does not bind down the scope of statistics to a particular field of enquiry.

Boddington denominates statistics as a 'science of estimates and probabilities'.[9] This is vague and indicates a narrow point of view. Estimates and probabilities are only a part and not the whole of statistics.

King defines statistics thus: 'The science of statistics is the method of judging collective nature or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates'.[10] The author himself regards it as possible that

---

[7] Bowley, A. L., *op. cit.*, page 3.
[8] Bowley, A. L., *op. cit.*, page 4.
[9] Boddington, A. L., *Statistics and their Application to Commerce*, page 7.
[10] King, W. I., *op. cit.*, page 23.

statistical problems such as would fall outside the limits of this definition might be imagined but maintains that it is sufficiently broad for practical purposes

In order that the definition of the science of statistics may suit its modern sense it should be so framed as to include all that is rightly its and to exclude what is extraneous to it Enough has already been said of what statistics and statistical methods are, yet a few observations are necessary before arriving at a suitable definition Statistics is concerned with mass phenomena, with large numbers descriptive of groups with results of collective action Individual facts and figures may be of interest to an individual Statistics does not deal with them For example the earnings of employees in a business may vary from man to man; a worker may earn Rs 7 in a certain week, or an average of Rs 5 8 0 per week and feel jubilant over it The business treats an employee's labour as a unit in the cost of production But an individual's income or a unit in the cost of production is not statistics nor does the gathering of knowledge about an individual's income become a statistical study If, however, we compare the total earnings of a group of workers with other elements in the business, say with its turnover, we arrive at a clearly defined relationship between them This relationship should hold good in normal circumstances. The study of this relationship is a statistical study Individual peculiarities count for nothing in it. It is the possession of the same peculiarities by the whole or a majority of the constituents of the group that is significant Further the data that are collected are influenced by a multitude of causes Statistics analyses them In studying the properties of such aggregates it employs methods that are based on particular characteristics of large numbers For instance a characteristic of large numbers and averages derived from them is that they enjoy great inertia individual incomes may chang

very fast, total or average income varies very little. Through such statistical methods accuracy of statements is examined, complicated data are analysed and one estimate is compared with another. All those estimates to which these methods apply fall within the scope of statistics. Statistics is, therefore, not confined to any particular branch of human knowledge: it is all-pervading. Theory of statistics should then comprise an exposition of statistical methods. It may, therefore, be concluded that **the science of statistics is a study of the methods applied in collecting, analysing, and interpreting quantitative data, affected by multiple causation, in any department of inquiry.**

Functions of a Statistician.

The functions of a statistician are then simple. He is concerned, firstly, with the collection of statistical data, secondly, with their analysis and finally, with the interpretation of the results of such analysis. Sometimes a sort of division of labour may be noticed in that the statistician may be engaged only on the analysis of data without bothering himself about the methods of collection or about the interpretations that may be put to his results. But such a division of duties may not always result in the best elucidation of a given problem.

A statistician cannot work miracles. He is not an alchemist expected to produce gold from any worthless material. He is rather like a chemist capable of assaying the value the material contains and of extracting nothing more than this value. It would, then, be no use commending a statistician because his results are precise nor condemning him because they are not. If he is gifted with the competence his craft demands, the value of his results shall follow solely from the material he analyses. His job is only to produce what the material contains, and no more. A necessary qualification of a statistician is that he must be an

impartial umpire free from fear or favour His personal preju dice should not be allowed to affect the conduct of his duties.

Main Divisions of Statistics

The domain of statistics can be generally classified into two main divisions Statistical Methods and Applied Statistics

Statistical Methods are concerned with the formulation of the general rules and principles applicable in handling different branches of data e g the methods of collection of data classi fication tabulation comparison by means of averages diagrams and coefficients correlation etc

Applied Statistics deals with the application of these rules and formulæ to concrete subject matter like wages prices trade, population Applied Statistics may consist of biometry, psycho metry vital statistics administrative social and economic statistics The last three are of immense importance and we shall be concerned generally with them

EXERCISES

(1) Explain clearly the concepts of Statistical methods and Statistical science

(2) What are the characteristics that statistical data possess? Explain them with illustrations

(3) Statistics are aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed enumerat ed or estimated according to reasonable standards of accuracy collected in a systematic manner for a predetermined purpose and placed in relation to each other

Elucidate the above definition bringing out clearly the characteristics of statistics.

(B Com Alld 1945)

(4) Comment on the following definitions of Statistics—

(*a*) By Theory of Statistics or, more briefly, statistics we mean the exposition of statistical methods.

(*b*) Statistics is the branch of scientific method which deals with the numerical aspects of aggregates of natural pheuomena.

(*c*) The theory of statistics comprises an analysis and interpretation of systematic collection of numbers relating to the enumeration of great classes.

(*d*) Statistics is that branch of science which deals with the frequency of occurrence of different kinds of things or with the frequency of occurrence of different attributes of things.

(*e*) Statistics is the science of estimates and probabilities.

(*f*) Statistics is the science of counting.

(5) 'Statistical methods include all those devices of analysis and synthesis by means of which statistics are scientifically collected and used to explain or describe phenomena either in their individual or related capacities.' (Secrist).

Elucidate the above statement.

(6) By statistical methods we mean methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes. (Yule)—Comment.

(7) What are the main functions of a statistician? Also point out the essential qualifications that one to be called statistician should possess.

(8) 'Statistics is the science of the measurement of the social organism, regarded as a whole, in all its manifestations'.

Sta t s s the science of count ng'.

Co t t on the above two statements and suggest any better d fin tion of stat stics giving reasons for preferring that defin tion

(B Com, Agra 1929)

(9) Comment on the followin g statement A statistician is not an alchemist expected to produce g ld from any worthless material

(B Com, All 1946)

CHAPTER III

FUNCTIONS AND IMPORTANCE OF STATISTICS

Functions of Statistics.

Statistics simplifies complexity. Human mind is unable to assimilate a mass of complicated data at any one moment. One can hardly form an unquestionable opinion regarding the comparative examination standards of two universities if he were simply told the marks obtained by every one of the two thousand students of each university. But if these unwieldly and complex data were simplified, reduced to totals or averages or presented through diagrams they would become readily intelligible. The Science of Statistics teaches how to do all this. It lays down methods which enable the boiling down of complex data to simple representative numbers easily adaptable to human mind.

Statistics enlarges individual experience. One may exercise his best ability and power of judgment to view the quantitative significance of a phenomenon. For instance, one may make a guess about India's national income at a particular time. But such a guess is subject to vagueness, inaccuracy and personal prejudices in the absence of adequate statistical data. And, when one proceeds to examine the accuracy of his statement he finds himself in the realm of statistical investigation. The resulting statistical estimate would be better than the conjecture of a casual observer.

Statistics compares the simplified data and measures their relationship. To appreciate the meaning of one estimate we often need another for comparison. A statement of water-rates charged in a certain town is meaningless if a similar statement for other equally important towns in the country is not forthcoming. It is, therefore, the relative or comparative, not so

much the absolute character of statistics that requires our atten tion But while making comparisons due allowance must be made for differences in the circumstances prevailing between two periods or two countries as the case may be For instance, if water-rates charged in a town whose source of water supply is situated at a considerable distance be compared with those charged in another town which has easy access to the source of water supply the result is bound to be vitiated if due allowance is not made for the differences in the conditions of the two towns

Importance of Statistics

Statistics has been termed as the Science of Kings Indeed, in ancient times it kept the kings informed about the man power and riches of their domain What are now called statis tical studies were in the past conducted under the name of *Political Arith metic* Civilisation has now advanced since then and the application of the mathematical theory of probability to social phenomena has yielded an **ingenious apparatus to deal with the figures of wealth and welfare** It will not be inappropriate, then to name statistics as the *Arithmetic of Human Welfare* to-day

Statistics is indispensable these days for a clearer appreciation of any problem affecting the welfare of mankind Problems relating to poverty unemployment food shortage protective tariffs, uneconomic agricultural holding etc cannot be fully weighed without the statistical balance These are the days of planning **Planning without statistics cannot be imagined** Neither can a policy be scientifically chalked out nor can its success be measured without the statistical apparatus

Statistics is the light bearer that enlightens the way to life s adventures It unravels the crowded complexities of life and thought Without its support man would wander aimlessly through this perplexing universe **Statistics discloses causal connection between related facts** Such study is at the bottom of all sound human endeavour

Statistics are the eyes of administration. No statesman can tender sound advice on a problem to his government unless he has adequate statistical data before him to base his judgment upon. Crime, drink-evil, tuberculosis and similar maladies need statistical investigation to suggest remedies for their cure. Budget, a collection of the estimates of revenue and expenditure of a state for the ensuing year, is an unavoidable necessity for an efficient running of government machinery. Its preparation is not possible without statistical records and without their utilisation by a personnel having knowledge of statistical methods. Once the budget is ready, decisions regarding enhancement or decrease in the existing rates of taxation or regarding the exploration of new sources of revenue can then be taken up without much trouble.

Statistics are an aid to supervision, particularly in these days of impersonal relationship between the employer and the employee. Every institution, commercial or otherwise, aims at obtaining efficiency with economy. Old plans are substituted by new ones. To test the effectiveness of new policies, the officers must be provided with accurate and consisely tabulated information showing the results obtained.

Statistics are invaluable in business and commerce. To be successful, a producer or a dealer should first estimate the demand for his wares, analyse the possible effects of factors like seasonal variations in demand, changes in taste, fashions and purchasing power of money, and then proceed to adjust his output or purchases to the estimates of demand. For, if he does not arm himself with a cautious and fairly accurate estimate, he would either be erring on the side of over-stocking himself and thereby suffering loss, or under-stocking himself and, therefore, losing chances of making profit. Business, indeed, runs on estimates and probabilities. The higher the degree of accuracy of a businessman's estimates, the greater is the success attending on his business. Correct estimation demands a high class of

skill which only long experience can ensure Statistics helps the recording of the past knowledge and experience, and drawing out standards or types with which results from year to year can be compared reasons for changes deduced and effects of such changes on future studied Experience so ascertained acts as an economic barometer The businessman forewarned of a currency inflation boom or depression shall prepare himself to face it boldly Statistics is closely associated with economic progress Statistics can be profitably employed in the different branches of commercial activity It is being applied in cost accounting The accounts of a concern undoubtedly show its financial position but they alone cannot correctly indicate business activity Statistical averages or indices shall have to be computed for reliable conclusions Statistics can help the launching of new projects and exploitation of potential markets In a word statistics is the life blood of successful commerce

An underwriter, a stock exchange broker or an investor in securities needs a knowledge of interest rates the fluctuations of investment market and other data to strike a timely bargain A banker, intending to build up a pyramid of credit should have an adequate knowledge of the seasonal variations in the calls for money on his bank to decide the amount of reserve that he should keep from time to time in his vaults

A railway operating over a wide area has numerous sources of possible wasteful expenditure Bad working such as using four engines where three will do or handling three tons where four should be disposed off is costly to the railway itself Similarly inability to clearly gauge the necessity of running special trains is not only to cause inconvenience to its patrons but also to deny to the railway the revenue that could have been its own if only conclusion had been drawn from past experience based on statistical records The rod of statistics is therefore indispensable for railway company to keep its working within the bounds of efficiency and economy

All forms of **insurance** subsist on precise calculations based on the analysis of a huge mass of data. The entire working of life assurance scheme, for instance, rests on the compilation of life tables and computation of expectation of life from time to time. Unemployment or sickness insurance similarly depends on statistical data.

Again, in order that a social or economic **legislation** may be fair and judicious it should be based on carefully recorded quantitative information. Enactments with regard to poor-relief, fixation of rate of exchange, levying of excess profits tax or stoppage of child marriage need a proper statistical investigation. The merit of the recommendations of various committees and commissions largely rests upon the statistical information behind them and the correctness of the statistical methods used.

Statistics is indispensable in a quantitative study. Its methods can be usefully employed in any science. A sociologist may attempt to demonstrate a relationship between sales of liquor and crime, or between suicide and poverty. A theoretical economist may make bricks out of the straw of statistics. He may deduce important economic principles from empirical data, or verify the validity of deductive laws of economics by the inductive method of statistics. Indeed, the science of economics is becoming statistical in its method. If one wants to study the march of time as revealed by the trend of population and the world's production of food, the relation of changes in the value of currency to prices, the relation of railway freight to internal and external trade the incidence of taxes, the influence of wages on health and efficiency of labour, the effects of irrigation etc., he must take recourse to statistical investigation. A serious danger in the absence of statistical informaion is to make random, arbitrary, estimates to suit one's ' pre-conceived ideas and pet notions. Statistics brings truth to light and corrects faulty observations. An economist should equip himself with

a knowledge of statistical methods to guard himself against possible fallacies of argument. Statistics has much furthered the development of economics and if the data are considerably large and reliable and correct statistical methods are used in collecting and analysing them even forecasts can be successfully made on their basis.

Statistical methods are extensively applicable. Astronomy pioneered their use in predicting about eclipses and biology has equally appreciably utilized them in its generalizations for instance in laws of variations and heredity. Meteorology uses them for weather forecasts. In these sciences and physics, geology, zoology etc. with whatever care and caution may the measurements be taken they cannot always be mathematically *exact*. And so an important problem to be attacked in them is to compute the most probable estimate—an average or a type—from a complex group of observations about which all the measurements are grouped in accordance with some *definite law*. The next task is to watch the nature and direction of the changes in the type or grouping of the measurements about it. Upon such study are based several generalizations and theories of these sciences which were they made arbitrarily and without statistical basis could not be fully relied upon *as true*.

Scope of the Science of Statistics

The above description of the importance of Statistics makes it clear that the science includes in its fold all quantitative analyses to whatever department of inquiry they might relate. Its scope is therefore stretched over all those branches of human knowledge in which a grasp of the significance of large numbers is looked for. Its methods provide an important manner of measuring numerical changes in complex groups and judging collective phenomena. Its scope is thus wide the limiting factor being its applicability to studies of quantitative character alone.

Limitations of Statistics.

It is necessary to carefully note the limitations of Statistics, otherwise one might make too much of this delicate science.

Statistics cannot take congnisance of individual items. It has already been emphasized that it deals with mass phenomena and as such throws light on the characteristics of the *whole* of the given group. Income *per capita* for a country, for instance, is nothing but the summation of its individual incomes divided by the total number of its inhabitants. Individual incomes differ widely, as is natural, from the poor to the rich. But these differences and irregularities of different items are brushed off in computing the average, which gives a graphic idea of the *per capita* income of the *entire* country. If this average is a fairly good figure, it by no means lessens the miseries of the under-fed, ill-clothed paupers and affords no reason why they should not be protected. The *per capita* income would not reveal the extent of their misery, much less that of the torture of each individual pauper. Consequently, where a study of the individual constituents of a group is desired other means should be restored to.

Again, **statistical results should not always be treated as the sole determinants of the value of a group.** Statistical method is one of the several methods of studying a problem. As such, it can be supposed to yield unquestionable inferences only when other available evidences and methods of study point to the same conclusions. Yet statistical results are as necessary for the quantitative appreciation of a problem as accurate measurements are for the construction of a building.

Besides, **statistical results might lead to fallacious conclusions if they are quoted shorn of their context.** If only the averages of the profits earned by two business houses over three years are quoted, and, incidentally, these averages

happen to be identical quantities one might conclude that both the houses were doing equally well It is not improbable that if one were to glance at the actual profits earned by them during the period the progress of one of them might be found to run in the positive direction while that of the other in the negative

Further **statistical laws are true on the average** or in the long run They are not like the exact laws of physical sciences which are said to hold true in every individual case that is subject to them Statistical laws therefore show approximate tendencies *e g* Pareto's law of income distribution

Not only that **statistical data must also be statistically uniform** That is the data should belong to a causal system that is highly stable so that there shall be no material fluctuations in its main characteristics over the whole field of observation Without homogeneity of data comparisons would be vitiated

Statistical methods are not applicable to the study of those facts that are not quantitatively measurable Health culture character friendship skill poverty cruelty and pessimism cannot be quantitatively expressed In studying any such phenomenon the statistical aspect is assumed to be subsidiary to other considerations A comparison of the state of civilisation between two countries does not lend itself to statistical treatment Resort may be had to such numerical data as the number of persons passing a certain standard examination the number of places of worship or entertainments or the number of people convicted of crime But these figures only indirectly relate to the real problem They are subsidiary to other information like the manner in which people in the two countries live the value they attach to principles of right conduct the treatment they mete out to others the type of work they perform the food they take and so on Therefore it follows that statistical methods are not of universal use or validity Their use is confined to quantitative studies

But **the greatest limitation of statistics is that only one who has an expert knowledge of statistical methods can scientifically handle statistical data,** since statistics, like medicines in the hands of quacks, are capable of being easily misused by the inexpert. One might harness statistics to his aid and make the worse appear to be the better case. Many people, therefore, look at figures with an eye of suspicion.

Distrust of Statistics.

There are said to be three degrees of comparison in lying: lies, damned lies, and statistics. One may not believe in the truth of a statement made without citing statistics. But, when he is presented with figures in its support he is led to believe—'if figures say so it can't be otherwise.' Such is the power statistics enjoy. And if this power is misused, say figures are deliberately manipulated, one may be, for the time being, led to take an utterly false statement for an absolutely accurate fact; but truth will be out some time later and when it is out figures that were cited in support of the statement would be labelled as lies. Cases where figures have been put forward as an evidence of the accuracy of a statement otherwise wrong, are not wanting and since lies and damned lies can be detected by a lay mind much more easily than a misuse of statistics, statistics have suffered the stigma of being classed with lies.

But reasons for disrepute cannot lie with statistics. By themselves they carry no weight. They can support false conclusions just as easily as they support the true ones. With them one may 'prove' that income *per capita* in India is high while others may 'prove' it to be low. What are these diametrically opposite conclusions due to? They are due either to motive or to ignorance. One may knowingly mis-handle the statistical apparatus with a will to prove his case, or ignorant of the characteristics of statistics and statistical methods run away with any figures, apply them

to any technique and feel happy over what he may regard his achievement Thus design and ignorance lead to misuse or abuse of statistics which in its turn leads to disrepute It is not statistics that are lies They are only tools in the hands of statisticians If tools are abused or misused it is not tools which are bad The fault lies with the way the tools are used

The popular distrust against statistics is generally expressed in the remark statistics can prove anything Those who say so are themselves at fault to a large extent As a matter of fact, little or nothing can be proved by statistics What can be done by them is to describe a phenomenon quantitatively classify it into parts summarize the facts relating to them and prepare the ground for a logical inference Very often too much faith is placed in figures alone and it is believed that the inference to which statistics lead is the *only* inference possible that, the inference is infallible and therefore need not be supplemented or verified by other than statistical evidence This is what should not be This has tended to bring the science of statistics and figurative data into discredit Conclusions must be made in part on evidence other than that offered by statistics

If figures are stated without their context, they are applied to a phenomenon other than the one to which they really relate, figures relating to a part of a group are given as relating to the whole figures favourable to an argument are stated omitting the other side they are inaccurately compiled deliberately manipulated and unscientifically interpreted—in all these cases they can be made to produce a false statistical argument All these apprehensions make many a man look at statistics with a jaundiced eye

Statistics suffer from the draw back that they do not always bear on their face the mark of their quality To a casual observer, a crude table compiled from unreliable data looks as valuable as

another prepared after great pains by a number of trained statisticians. Most often, people who are served with statistical information do not know whether a particular factor can be statistically evalued, or whether the information is based on satisfactory data. If they are men who believe that 'figures won't lie' they shall accept them without question, while others who are sceptical of their truth shall treat them as 'tissues of falsehood.' In fact, before accepting or rejecting statistics, one should enquire into the competence of their author. Another apparent draw-back of statistics is that since they are expressed as definite, concrete quantities they look innocent and precise, and people often believe them to depict an accurate picture. But if they are disillusioned they blame them. It should be noted that their appearance in quantitative form is not a guarantee of their accurately presenting the phenomenon to which they relate. They show only one method of doing so.

Whatever be the distrust of statistics it does not imply that statistics have no value, or that the Science of Statistics is useless Drugs may be misused, but for that reason neither is their usefulness lost, nor does the science of medicine become valueless. No doubt statistics do not supply conclusions but they do furnish, in part, the basis on which conclusions may be drawn. Their usefulness is, therefore, great. It is imperative then, that statistical data should be handled only by those who are aware of their use, limitations and dangers and are free from prejudice. If their limitations are forgotten fallacious conclusions would result. If data are carefully collected and scientifically analysed, the results obtained shall be trustworthy. With the study of Statistics as a science, with the recognition of its imitations and with improvements in its technique the cause for its distrust is gradually waning. A layman is apprehensive of statistics largely because he does not understand the technique which statisticians apply to a problem in which he is interested. If he learns to exercise his

discriminating mind and critical faculty to judge for himself before blindly accepting or summarily rejecting any statistics that he is served with the misusers of statistics will find it hard to ply their trade

EXERCISES

(1) Explain and illustrate the functions of statistics

(2) Discuss fully the importance of statistics as an aid to commerce

(B Com Alld 1942)

(3) Statistical methods are most dangerous tools in the hands of the inexpert *Statistics is one of those sciences whose adepts must exercise the self restraint of an artist*

Explain fully the significance of the above statement

(M A Patna 1942)
(B Com Alld 1947)

(4) Write an Essay on—
Either (a) Statistics in the service of the State, or (b) Collection of economic statistics during a population census

(Dip in Econ Madras 1931)

(5) A knowledge of statistics is like a knowledge of foreign language or of algebra it may prove of use at any time under any circumstances —Explain

(6) Give the important uses and limitations of Statistics Show its relation to Economics and Mathematics

(B Com Luck 1938)

(7) Discuss the utility and value of statistics in the measurement of social phenomena How far do you agree with the view that economic planning without a proper organization of statistics of the country is inconceivable?

(M A Agra 1940)

(8) The Statistics of a business can be treated scientifically and the preparation and study of business statistics may be made a *more exact science than the study of national and social statistics* Explain

(B Com, Alld 1932)

(9) Explain clearly the statistical methods used in any scientific investigation, and show their importance to theoretical economists and practical businessmen.

(B.Com. Alld., 1933).

(10) Give a lucid explanation of limitations of statistics.

(11) "There are three degrees of comparison in lying. There are lies, there are damned lies, and there are statistics"—How far do you agree with this statement?

(12) How do you reconcile the following statements?—

I. (*i*) Statistics can prove anything.
(*ii*) Statistics do not prove anything.

II. (*i*) Statistics are lies.
(*ii*) Figures do not lie.

(13) Discuss the scope, utility and limitations of statistics.

(B.Com., Agra, 1937).

(14) The claims of statistics to our support depend upon the efficient mental training it provides for the citizens, the light it brings to bear upon many important social problems, and increased comfort it adds to practical life—Discuss.

(15) Describe briefly the different kinds of statistical methods, and show the usefulness of the knowledge of statistics to businessmen, scientists, and economists.

(B. Com., Agra, 1936).

(16) Explain clearly what *you* understand by the Science of Statistics. Discuss its scope and limitations.

(B. Com., Alld., 1944).

(17) 'Sciences without statistics bear no fruit, statistics without sciences have no root.'

Explain the above statement with necessary comments.

(M.A., Patna, 1943).

(18) In what ways can statistical methods be misused by interested persons? Give at least two examples of the misuse of statistics

(B Com Luck 1939)

(19) Discuss the importance of the study of statistics and show how it can help the extension of scientific knowledge the establishment of a sound business and the introduction of social and political reforms

(B Com. Agra 1942)

(20) Statistics are like clay of which you can make a God or a Devil as you please Discuss

(B Com., Alld 1948)

CHAPTER IV

STATISTICAL INQUIRIES AND UNITS

Collection of figures is an all-important work in a statistical study. The actual collection of data must be preceded by a well-thought out plan. In planning a statistical inquiry it is necessary to

(1) ascertain the purpose of the inquiry;
(2) determine the type of inquiry to be conducted in order to gauge its scope and the available sources of information;
(3) define the statistical units in terms of which data will be collected and compared;
(4) lay down the degree of accuracy desired; and,
(5) decide upon the method of collection of data.

Purpose of Statistical Inquiry.

In organising a statistical inquiry it is at first essential to ascertain the object of the inquiry, since the type of inquiry to be undertaken and its details will be largely determined by the light which it is the purpose of the inquiry to throw. The purpose of an inquiry, for example, may be to study the cost of living of industrial labour in a town during a certain period.

Types of Statistical Inquiries.

An appreciation of the different types of statistical inquiries is necessary for, the meaning, scope and accuracy of statistical data and the method of collecting requisite information are dependent upon the type of inquiry in hand. Distinction between statistical inquiries can be made upon answer to the question—**By whom is statistical information required?** It may be

required by the state a business house or a scientific investigator Their facilities for collection of data differ The state may legislate, an institution may request while a private individual may have to beg for the purpose The sum of money that everyone of them can spend on the inquiry is different, individual s financial capacity being the weakest Again official commercial and scientific inquiries will look at the same subject matter differently, facts material to one class of investigation may not be relevant to another

Another distinction between statistical inquiries can be made according to **how the statistical information emerges.** Figurative data may emerge as by products of certain administrative operation or they may be obtained directly by collecting information relating to certain affairs In the first case collection of data is not the primary purpose, but subsidiary to or only a part of the main action For instance, imports into India are recorded at the Customs Office and these records serve as the raw materials of statistical tables In the second case collection of data is the sole end. For example census of population yields the figures which it is the purpose of the census to collect Obviously, the degree of accuracy attainable in the results of the first type of inquiry is not likely to be higher than that in the results of the second that is an inquiry *ad hoc* Again, in the first case, the definitions of the terms used shall be so designed as to suit administrative needs while in the second they shall suit the purpose of the particular problem For instance, the term 'wage may mean money wage to those administering sickness insurance scheme and 'real wage' to wage earners claiming dearness allowance in times of rising prices

Again statistical inquiries may be distinguished according to the **source from which the information is obtained** In some inquiries the number of persons playing an influential part may be large, in others small For administering a scheme of food

rationing in a town, every householder may be held responsible for supplying information relating to persons and grain-consuming animals in his household. The number of house-holders is, no doubt, very large. Therefore, the sources from which information is obtained are varied. The questions contained in the form will, therefore, have to be few, simple and unambiguous because of the varying educational standards of the people. The scope of the inquiry would naturally be limited on this account. If, on the other hand, sources are few compared to the size of the inquiry, these few may be skilled investigators appointed to collect requisite information. The scope of the inquiry can be extended here, because the investigator can elicit the information which, in the former case, may be difficult to extract.

Another distinction between different kinds of statistical inquiries may be made according as the inquiry is of **census** type or a **sample** survey. In a census the whole group is surveyed as, for instance, the Census of Population or Accounts of Foreign Trade of India. In a sample inquiry only a part of the group is surveyed as, for example, a sample survey of acreage under jute in a Province.

Statistical inquiries may also be distinguished as **direct** and **indirect.** Height of students in a class is measurable directly in inches, but their intelligence cannot be quantitatively ascertained. In cases where the desired information is not capable of statistical treatment, some allied information reducible to numerical standards will have to be collected. In this particular case reliance will be placed on the marks obtained at a certain examination or intelligence test. This inquiry is indirect.

Statistical inquiries may be **original** or **repetitive.** They may be carried on either for the first time, or in continuation or repetition of previous inquiries. In the former case a plan will be initiated. In the latter, old plan with such minor alterations as experience or necessity demands may be followed. But the

definition of units used should not be materially altered in the repetitive inquiry Advantages resulting from modifying the old plan and from continuity and comparability of information must be weighed before affecting any alteration in the plan

Lastly statistical inquiries may be undertaken for absolutely confidential purposes or they may be thrown open to public. Trade associations may collect information from their members which may be kept secret Modes of treatment for both types of inquiries will not be identical

When decision has been taken on the type of inquiry that would be undertaken the scope of the inquiry would become clearly visible and the available sources of information should be carefully studied

Units of Measurement.

Having ascertained the purpose for which statistical data are to be collected and used and having formulated the type of which the inquiry will be, the next step in organizing statistical studies is to define rigidly and unmistakably the units of measurement in which the aggregates to be counted shall be expressed Quantitative science demands a precise and unambiguous terminology for the terminology once specified shall be adhered to throughout the inquiry Adherence to the definition once made is essential in order that the thing counted or measured may be the same throughout the inquiry Strict comparison shall be possible only when the things counted are the same The task of defining the unit seems at first easy but in many cases the opposite is true Literacy connotes one meaning to an ordinary person another to a sociologist but for understanding the tables relating to it in the Indian Census Reports its meaning is something precise —ability to write a letter and to read the answer to it In studying the problem of educated unemployed in India the questions that at once arise are 'what is exactly understood by educated ? and

who is 'unemployed'? Upon a little thought it will be found that it is not easy to answer such simple questions. Similarly, factors like wages, profits, accidents, imports, investments are differently interpreted by different people. *Correct definition is always determined by the purpose in mind.* Different purposes will necessitate different definitions of the *same* unit. But before collection of data begins a correct specification of the unit will have to be made.

A statistical unit should satisfy the following requirements:—

1. The unit must suit the purpose of the inquiry.
2. Its definition must be unambiguous, simple and complete in itself.
3. The unit must be definite, specified and ascertainable,
4. The unit must be stable and standard. (In India, currency fluctuations have not been rare, and weights and measures still vary from locality to locality. Hence the necessity of taking stable and standard unit.)
5. Homogeneity and uniformity must be ensured. A unit should not imply different characteristics at different times, If the data are heterogeneous, they may be broken up into small classes to secure uniformity, or the process of standardization may be followed. For example, the data for the compilation of an average of the wages received by workers in a factory, where male and female adults and children are working side by side, are heterogeneous, women getting lower wages than the men, and children getting the least. In order that the average wage may be a true representative, the data may either be sub-divided into three group, *viz.*, 'wages for male adults', 'wages for female adults' and 'wages for children,' or females and children may be expressed in terms of equivalent men.

The units of measurement may be classified into:—

I. Units of Enumeration, and

II. Units of Analysis and Interpretation.

Units of Enumeration Simple and Composite

Un ts of enumeration are those in which measuren ents are made They are therefore concerned with the collection of data These units are of two kinds simple and composite

A **simple unit** is one that denotes a combination of character istics that occur together It simply dist ngu shes classes Its meaning is general Examples of simple units are a ton a pas senger an accident a sale a store a house a room Such units are mutually exclusive They are defined easily and fairly precisely The degree of error associated with them is therefore small

A **composite unit** is one that is formed by adding a limiting or qualifying word or phrase to a simple unit with the result that its scope becomes limited and the task of defining it does not remain easy Examples of composite units are ton mile passenger mile man hours

Units of Analysis and Interpretation

These units are those in terms of which data are compared. They include such units as rates ratios and coefficients They will be discussed in Chapter IX It may however be well to note here what a coefficient means

A **co-efficient** takes the form of comparative statement. Comparison may relate to time to space and to conditions in time or space For example wages may be expressed in Rupees but related to days or months We then speak of Rupees per day or per month. We may express production of wheat in maunds but relate it to province farm or acre We then speak of maunds per farm or acre or province. We may lastly express income of a given area in units of money *but relate it to the* entire population. Then we speak of income *per capita* A co-eff cient in effect is a comparison between the numerator and the denom nator both of which should be related and homo-

geneous. If passenger-miles are divided by passenger-train-miles we shall obtain passengers per train. But if passenger-miles are divided by ton-miles a monster will result.

Degree of Accuracy.

After the units have been selected and defined, the next step in planning a statistical inquiry is to lay down the standard of accuracy to which the collected data should conform. It should be decided whether, for instance, prices shall be quoted correct to a pie, or to an anna, or to a Rupee. Every phase of questions should be carefully studied and details checked. No effort should be spared to minimise the chances of error. The aim should be to avoid the necessity of conducting a second inquiry and thus save time, labour and money from being wasted.

Then, as the last step in organising an inquiry, a suitable method of collecting statistical data will have to be selected.

EXERCISES

(1) Explain with examples the different types of statistical inquiries, and indicate the bearing of each on the collection of data.

(2) What do you understand by the 'object' of an inquiry? Is it necessary to determine it before planning a statistical inquiry? Why?

(3) What are statistical units of measurement? Explain the necessity of determining them.

(4) Differentiate between simple and composite units. Give illustrations of transforming units from simple to composite.

(5) What difficulties are experienced in defining the following terms for collecting statistical data?

Accident, Industrial-accident, Room, Class-room, Hindu, Exports, Literacy, Book, Improved variety of crop, Wage.

(6) Differentiate with examples the units of measurement from the units of comparison

(7) What is a co-efficient? Illustrate with suitable examples

(8) What precautions should be observed in specifying a unit?

(9) What preliminary steps will you take in planning a statistical inquiry?

(10) What is a statistical unit? Is it necessary that the data be homogeneous?

(B Com Agra 1939)

CHAPTER V

COLLECTION OF STATISTICAL DATA

Primary and Secondary Data.

Statistical data are generally classified as primary and secondary. The former are those which form the raw material of inquiry, while the latter are those which have gone through the statistical machine at least once. The former are original, *i.e.*, those in which instances have been recorded as they occurred without having been grounded at all. The latter are those that have been worked up to a certain extent, *i.e.*, they have been collected, tabulated and presented in some suitable form for any purpose. They are generally expressed in totals, averages and percentages.

Distinction between primary and secondary data is one of degree. Data which are secondary in the hands of one may become primary in those of another. For instance, statistics of foreign trade of India are secondary data to the general public while they are primary data in the hands of the statisticians of the Department of Commercial Intelligence and Statistics. The distinction between them lies in the fact that when figures have been 'worked over' for a purpose, when they have been examined for their accuracy and comparability and have been grouped, averaged or reduced to percentages—that is, when they have lost their individual characteristics which they possessed when they were reported—they become secondary data.

On the basis of primary and secondary data the methods of collecting statistical material have been divided into Primary Method and Secondary Method, the former being used in the collection of original data, while the latter in the collection of such data as have already been 'worked over' to some extent.

Primary Method

Under this method the following ways of collecting the requisite data are generally used.

1 **Direct Personal Observation.**—This method yields very accurate results for it implies that the investigator must be present on the spot to make patient and careful personal observation regarding how people work and live. First hand information collected thus must be reliable. But since within a reasonable amount of time an extensive field of inquiry cannot be covered by this method it is useful specially for intensive studies It is therefore utilized in localized inquiries Besides covering only a narrow field this method is open to the charge that the chances of personal prejudices of the investigator affecting even unconsciously his conclusions are great

If it is not practicable for the investigator to be on the spot or to devote the time the above method demands an alternative for the investigator is to question and cross-examine a person who is directly in touch with the facts under investigation Here since the investigator counts on the goodwill of others he will have to be courteous in his behaviour Besides the questions that he would ask must be few very simple clearly worded not inquisitorial and so far as possible demanding an answer in yes or no or a number This alternative method is also used in intensive studies

2 **Indirect Oral Investigation** assisted by a standard list of questions —When the information desired is complex and informants are indifferent to supply it if directly approached or if the field to be covered is very extensive so that the first method cannot be successfully employed the *viva voce* indirect evidence of several third parties preferably of those indirectly in touch with the facts under inquiry may be recorded Commissions and Enquiry Committees appointed by governments generally find this method suited to their needs But certain

precautions must be observed in order that reliance may be placed on the data collected. Firstly, the indirect evidence of one person should not be relied upon. Secondly, it should be clearly ascertained whether the informant really possesses a knowledge of full facts. Thirdly, it should be considered whether the person questioned is prejudiced in favour of or against a particular viewpoint or is motivated to colour the facts. Due allowance must be made for the optimism and pessimism of the informant. Lastly, if the informant happens to be an uneducated person or suffering from occasional fits of mental disequilibrium, it should be seen whether he would be in a position to give expression to his ideas adequately and precisely.

3. **Estimates from Local Sources or Correspondents.**—This method does not imply a formal collection of data. Local correspondents obtain the estimates in their own manner and report them to an appointed authority. This method yields only approximate results, but expeditiously, at a small cost and with ease.

4. **Investigation through schedules to be filled by the informants.**—This method differs from the preceding one in that the questions asked of the informants are those in respect of which they are supposed to have definite and precise information. If the informants reply intelligently, this method is good for extensive inquiries. It is comparatively less expensive and fairly expeditious. This plan is largely adopted by private individuals and even the government. But a large number of informants do not generally answer the schedules unless the inquiry is in the informants' own interest, or the private individual or institution responsible for the inqury is able to persuade them to answer, or the state exercises its legislative powers for the purpose. And, schedules that are returned are very often incomplete, ambiguous and full of errors, since the average informant is indifferent in these matters. In order that correct answers may be had, the schedule,

or a letter of request attached to it, should state the purpose of the inquiry and the identity of the person or the institution responsible for the inquiry, and give an assurance to treat the information tendered as confidential if so desired by the informant. The questions should be few, clearly phrased and, above all simple and should not be such as to arouse suspicion and prejudice in the informant

5 **Investigation through schedules in charge of Enumerators.**—Under this plan the informants themselves are not to fill in the schedules Instead the trained enumerators put their questions and record their answers Therefore in this kind of inquiry the schedules can be much more exhaustive than those in the previous one and the scope of inquiry can also be enlarged Correct interpretation of every question and the method of collecting informations must be explained in detail to the enumerators, so that different enumerators may not give different weight or meaning to the same questions Enumerators should be equipped with a sample schedule duly filled in. This plan affords quite good results and is the best for many extensive investigations It is generally adopted in large-scale governmental inquiries, Its cost is no doubt prohibitive to a private individual or institution

Choice of Enumerators.

Selection of enumerators should be done with great care since on them depends the quality of the investigation Intelligence diligence and integrity must be their attributes in order that vague replies of informants may be detected corrected or eliminated and fictitious quantities may not be entered. The enumerators should also be courteous and tactful so that they may extract the requisite information without causing resentment or ill will in the informants They should be free from bias If these precautions are taken in the selection of enumerators needless errors and confusion shall be avoided

Choice of Questions.

A word about schedules used is necessary. The schedules may be either what are called 'Questionnaires,' the answers to which are recorded on a separate piece of paper, or 'Blank Forms,' in which space is provided in the form itself for filling in the reply. Headings and titles drawn up in them should be lucid and easily understandable, and the degree to which accuracy of a numerical result is required should be indicated. The size of the paper should not be unwieldly, and each word and phrase used should be carefully scrutinized for ambiguous or controversial interpretation.

The questions which are asked should be—

(1) such as the informant shall be able to answer,
(2) few in number,
(3) simple and clear enough to be easily grasped,
(4) not inquisitorial and not causing resentment so far as possible,
(5) requiring brief answer, say 'yes', 'no' or 'a number,'
(6) corroboratory, if possible,
(7) capable of being answered without prejudice, and
(8) directly related to the point of information desired.

Selection of Representative Data.

When the inquiry in hand is very extensive, it will not be practicable to undertake a census type of inquiry where each individual item of the universe or 'population' shall be questioned. The inquiry will have to be of the type of sample survey, and the sample will be representative of the whole field.

The methods of selecting a sample or representative data are two :—

1. Deliberate or Purposive Selection,
2. Random Sampling or Chance Selection.

Deliberate Selection.—According to this method a small group out of the given universe is selected at the discretion of the organiser of the inquiry the guiding principle being that the selected group should be representative of the given universe so that its study could reveal the characteristics of the mass If economic conditions of people living in a province are to be studied according to this method a few towns and villages and also the required number of families therein may be deliberately selected for intensive study on the principle that they shall be typical of the entire province But the fear is that they may not always be typical since *personal element* has *a great chance of entering* into the selection of the sample The organiser may select a sample which would yield results favourable to his point of view and the entire inquiry may thus be vitiated. If the organiser and investigators are unbiassed the results obtained from an analysis of deliberately selected sample may be tolerably reliable provided the basis of selection is otherwise unquestionable

Random Sampling.—Chance selection on the other hand is free from the influence of the personal factor It is so to say a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process so that every unit has equal probability of entering the sample. Here it is blind chance alone that determines whether the one unit or another is selected This is an important condition of this method If from a list of villages of a given area arranged in alphabetical order every 100th or 50th village is marked out for intensive study it would give a hundredth or a fiftieth sample of the whole area Or in urban areas from a directory of the owners of shops in a particular locality every 10th or 20th shop may be selected for intensive study This type of random sample surveys for rural and urban areas were suggested by the Bowley Robertson Committee for India It should however be noted that once a random selection of villages and shop has been made on no account should any one of the villages and shops so selected be substituted by another

Census *vs.* Sample.—In a census inquiry we may, for example, establish the facts about the heights of 2,000 people by finding averages and other statistical indices: our problem shall, then, be limited to a characterization of the heights of these 2,000 people. But, in a sample survey we shall use the properties of a random sample of variables for drawing inferences about the larger population from which the sample is drawn. In the above example, our question will be: what approximate or probable inferences can be drawn regarding the stature of a whole race of people from an analysis of the heights of a sample of 2,000 people drawn at random from the people belonging to that race? Thus, in a census inferences for the whole population are drawn from a study of the whole field, while in a sample survey inferences for the whole population are drawn from the study of a representative part.

The 'Sample' method has many advantages over the 'Census' method. It economises in time, labour and money and permits a small band of skilled investigators to do the whole job more efficiently and precisely than a large army of unwilling, inefficient enumerators would do in a census. Random sample survey is largely coming into use because of these advantages and of its scientific character. It affords a sufficiently accurate picture of a large group without requiring a resort to a complete enumeration of all the units of the group.

Theory of Probability.—The method of random sampling is based on the mathematical 'Theory of Probability.' The theory implies that if from a very large group of items, technically called the 'population', a moderately large number of items is chosen *at random* such numbers are almost sure, *on the whole*, to possess the characteristics of the population. If two men plucked 200 leaves each of a particular tree, the averages of the lengths of the two sets of leaves would be almost identical, even though the leaves may vary considerably in size. Further, if one were to obtain the average length of all the leaves of the tree, it

would not materially differ from the average length of either group Similarly, if a rupee coin is tossed twelve times, the probability is that it will fall half times (i.e., six times) with its head or tail turned upwards On this principle gamblers—dice-throwers, card players, etc.,—run risks continuously and with profit, and insurance companies insure against death or other calamities. It is this principle that is responsible for regularity in the number of crime and suicides in a country for a given period of time This principle is christened the **Law of Statistical Regularity.**

It should, however, be noted that *any* number of samples will not give exactly the same results as a study of the whole group would As a matter of fact, the probability of error diminishes with an increase in the number of items included in the sample. That is, the larger the sample, the more reliable are its indications. Its reliability is proportionate to the square root of the number of items included

Law of Inertia of Large Numbers.—This law furnishes a corollary to the law of statistical regularity It implies that large numbers are relatively more stable than small ones. If one part of a large group varies in one direction, the probability is that another equal part of the same group would vary in the opposite direction so that the total change would be slight For example, the production of tea or wheat may vary from locality to locality in a given year, but the total world production of tea or wheat remains relatively stable for decades Deaths in different parts of a country in a given year may show violent fluctuations, but all fluctuations will hardly be in the same direction, so that death rate for the whole country will remain almost constant through a number of years Thus large numbers and the averages deduced from them have great inertia.

But it does not mean that the property of inertia does not allow for change with the passage of time It only signifies that

if the set of figures under consideration are very many the change is likely to be more regular than it would be when the figures considered are few. Secular movements resulting from long period tendencies in the background of conditions are not precluded. The death-rate of a country, for example, may be relatively stable from year to year, but for a long period it may show a progressive decline.

Secondary Method.

Data already collected by others may be in manuscript form, for example, original records of (*a*) business houses or of (*b*) government offices, such as accounts of business firms and public offices, patwari's village-books, etc., or (*c*) notes of past investigators and chroniclers. They may be type-written or printed matter. Printed documents include books, journals, reports, bulletins and official publications. They may be published or meant for private circulation only. They may be original or derivative. They may be by-products of administration or meant for public information. They may be official or non-official.

The following different ways of compiling secondary data for statistical inquiries may be noted:—

1. **Utilizing published information.**—Such information may be—

(*a*) official, *i.e.*, published by Government,

(*b*) semi-official, *e.g.*, published by municipalities, railways, etc.,

(*c*) published by technical and trade journals,

(*d*) published by trade association, chambers of commerce, etc.,

(*e*) published by research institutions such as universities, Economic Enquiry Boards,

(*f*) published by individual research workers.

2 **Utilization of Business Intelligence Service bulletins,** *e g* daily weekly or monthly bulletins market reports issued by Stock Exchange or Produce Exchange and dealers of repute and standing

3 **Utilization of unpublished data or manuscripts.**

4 **Utilizing information collected by other agencies or for other purposes.**

EXERCISES

(1) How will you organize an economic survey of a small Indian State comprising five towns and one thousand villages?

(M Com. Alld. 1943)

(2) How far do the results of statistical investigation depend upon correct sampling? Compare the different methods used to secure representative data

(B Com Agra 1939)

(3) Explain in detail how you would proceed to organize a Census of Wages Draw up a blank form or forms to obtain the information required

(B Com Agra 1937)

(4) Compare the advantages and disadvantages of the Census method (or complete enumeration) and the Sample method of collecting statistics

(B Com Cal 1937)

(5) Mention the different kinds of statistical methods generally used in investigations. Are there any fields of inquiry where these methods cannot be used satisfactorily?

(B Com. Agra 1940)

(6) What is a Questionnaires? How does it differ from a Blank Form? What precautions should be taken in drafting questionnaire?

(7) Draft suitable questionnaires for enquiries regarding:

(1) Educated unemployment in India.

(2) Sugar industry in U. P.

(3) Economic condition of agriculturists in India.

(4) Cost of living of a University staff.

(5) Budgets of students in a college.

(6) Industrial survey of U. P.

(8) How would you organise an investigation into the hand-loom weaving industry of the U. P.? Prepare questionnaire suitable for the purpose.

(B.Com., Alld., 1942).

(9) It is required to estimate the total consumption of food-grains in the U. P. for enforcing a scheme of food-rationing. What statistical data should be collected for the purpose, and how?

(10) Show the necessity of the use of the method of *random sampling* in any extensive investigation. How woull you make use of this method in carrying out an economic survey of the rural areas of U.P.?

(B.Com., Alld., 1935).

(11) Statistical investigations carried out by the Govt. are usually based either on complete enumeration of the universe of reference as, for instance, the population census, or on the study of 'typical' cases as, for instance, the proposals regarding the economic census. Explain why the method of random samples is to be preferred to either of these methods.

(M.A., Alld., 1935).

(12) Write brief, but lucid notes on

(a) Law of Statistical Regularity

(b) Law of Inertia of Large numbers

(c) Primary and Secondary methods of collecting data

(13) Fecundity among the poor is higher than among the rich This statement is often made It means that poorer classes of persons beget more children than richer ones How would you verify statistically the truth or otherwise of this statement? Describe fully the method you would employ giving examples

(B Com Agra 1934)

(14) Explain the whole process of organising an enquiry into the system of agricultural marketing in the U P

(15) A cotton manufacturer in Bombay is anxious to find new markets for his goods in India and foreign countries What statistical materials should he collect? What material would he be able to get from published documents?

(B Com Alld 1935)

(16) A sugar manufacturer in the U P is anxious to find new markets for his sugar outside India Describe the procedure he should follow to get all the necessary statistical information for the success of his mission.

(B Com Luck. 1938)

(17) State and explain the law of Statistical Regularity Discuss the methods generally used in sampling

(B Com Agra 1941)

(18) Classify the methods generally employed in the collection of statistical data, and state briefly their respective merits and demerits.

(B.Com., Alld., 1946).

CHAPTER VI

EDITING THE COLLECTED DATA

Editing Primary Data.

After the schedules have been returned by the informants or enumerators as the case may be they should be edited, i e scrutinized to detect errors omissions and inconsistencies If possible, defective schedules should be sent back for correction or if the investigator has reasonable ground to do so, he may himself make the required amendments Undue tampering with them is however dangerous Only in cases of unmistakable error should alterations or modifications be made otherwise even with a will to be impartial a wrong fallacious conclusion might result Such schedules as are thoroughly unsatisfactory must be rejected For a smaller number of correct samples is better than a large number of incorrect ones In the former case the error can be mathematically corrected with approximate accuracy It is not possible to do so in the latter If majority of informants have misunderstood a question there is a clear case for making a change and conducting a second inquiry The extent to which omissions may be allowed is also of importance. If the returns unquestionably confirm a certain fact and the samples are tolerably representative the omission of a number of returns does not matter If on the other hand evidence is conflicting the omission of even one return may be a serious matter The degree to which lack of accuracy presence of errors or approximation is to be tolerated in editing the data is of great significance

Accuracy.

Perfect accuracy in the data is rarely attainable Wheat crop in India for instance cannot be exactly measured It can

be estimated to a reasonable degree of accuracy. A weighman, however perfect his weighing balance, cannot weight wheat or any other commodity correct to within, say, 1/64th of a seer. Similarly, the distance for a four-mile cross-country race cannot be measured without giving a probable error of a few yards. Even in scientific measurements absolute accuracy is unattainable, for, heights of liquids in test-tubes may vary by a thousandth part of an inch or angles may differ by a hundredth part of a degree. Thus, absolute exactitude is not possible; a closer approach to it is. Reasons for it are obvious: the observer and the observing instruments are both sources of error; statistical data cannot be given hard and fast definitions; the statistician, unlike a chemist, cannot experiment, conditions being outside his control. Statistical methods do not, therefore, aim at arithmetical precision A statistician would be satisfied with *reasonably* accurate figures provided he can measure their reliability. To attempt to obtain the greatest possible degree of accuracy is to waste time. Statistics has, thus, to deal with estimates and probabilities and not exact enumerations.

What is a reasonable degree of accuracy shall be determined by the nature of the material and the purpose of its measurement. In common use, only a certain conventional accuracy is required. Precious metals or drugs are much more minutely weighed than hay, husk or saw-dust. Height of a room may be measured correct to an inch, but the difference of half an inch in the length of a man's nose will make him a monster. A railway administration is satisfied if a parcel for booking is weighed correct to a seer, but the post-office would weigh it correct to a tola at least. It would be enough if we can measure an object to that degree of accuracy which is required for practical purposes. Thus **it is relative and not absolute accuracy that is desired in statistical data.**

Statistical Errors

The word error is used in a special sense in statistics. It denotes the difference between the true value and the estimate of a quantity. It differs from a mistake in that it refers to difference resulting from any source of inexactitude.

The chief sources of errors are three. First, **errors of origin**, *e.g.* errors resulting from a prejudiced information or inappropriate definition of units; second, **errors of inadequacy**, *e.g.* errors due to inadequate sample data or incomplete information; third, **errors of manipulation**, *e.g.* unconscious error in measuring, weighing, counting or approximation.

Measurement of Error

Errors may be measured as absolute or relative. An investigator is concerned with relative error more than the absolute one. **Absolute error is the difference between the true value and estimate of a quantity, while relative error is the ratio of the absolute error to the estimate.** If the monthly average expenditure of students in a hostel is in reality Rs. 50 and we measure it as Rs 49, the absolute error is Rs (50—49), *i.e.* one rupee while the relative error is

$$\text{Rs. } \frac{(50-49)}{49} = \frac{1}{49} = \text{Rs } 0.0204$$

The relative error is sometimes expressed as a **percentage error**. Thus the relative error in the above case is 2.04 per cent. The error is positive since the true value exceeds the estimate. If on the other hand we estimate the monthly average expenditure as Rs 51, the absolute error is Rs. (50—51), *i.e.* minus one rupee and the relative error is

$$\text{Rs } \frac{50-51}{51} = \text{Rs } \frac{-1}{51} = \text{Rs } -0.0196$$

The percentage error is

$$\frac{-1}{51} \times 100 = -1.96$$

The error is negative since the true value is less than the estimate.

In algebraic notation, let u represent the measurement of a quantity whose true value is u^1, and e stand for relative error of the estimate, then

$$e = \frac{u^1 - u}{u}$$

and, if ue stands for absolute error,

$$ue = u^1 - u$$

The error would be positive or negative according as u^1 is greater or smaller than u.

Biassed and Unbiassed Errors.

Errors may also be classified as biassed and unbiassed. **Biassed errors** result from a bias or prejudice on the part of the informant, enumerator or measuring instrument. These errors, therefore, lie in the same direction and are cumulative in character. That is, the greater the number of observations, the greater would be the absolute error. For example, if wall-paper were measured with a foot-scale half an inch short, then the greater the number of feet measured the greater would be the absolute error.

Unbiassed errors are those which arise automatically, without any bias or prejudice. They are subject to the law of statistical regularity, so that the excess in one direction is almost balanced by the defect in the other. Unbiassed errors, therefore, are compensating. The larger the number of items, the smaller will be the absolute error. If the foot-scale in our example be correct, error in a measurement in one direction shall be compensated by error in another measurement in the opposite direction, so that the greater the number of feet measured, the smaller will be the difference between actual length and the length measured by the scale.

If some investigators carry on investigation into the economic condition of agriculturists in a few places in, say, the U P in India pre-determined to prove that their income is high they would probably by examining only the well-to-do and debt free cultivators and taking the incomes of those living near the grain markets having their own pack animals for transporting grain and marketing their produce themselves, produce a high average income for each locality But if they are not prejudiced by a pre determined conclusion, that is if the inquiry is impartial the investigators are as likely to make a low estimate in one locality as to make a high one in another In the former case, errors are biassed all being in the same direction and causing the average to go up In the latter case, errors are unbiassed, positive ones neutralizing the negative ones and reducing the resulting errors to a small figure The following illustration would clear the point —

Average monthly income in	Fact	Biassed Estimate	Unbiassed Estimate
	Rs	Rs	Rs
Locality A	20	21.5	21
Locality B	16	17.5	15
Locality C	22	22.6	22
Locality D	18	18.4	18.8
Averages	19	20	19.2
Relative errors		5%	1%

From the above table it is clear that the errors of biassed estimate are cumulative, while those of the unbiassed one are compensating or counterbalancing Another illustration of biassed and unbiassed errors is provided by the age-returns in the Indian Census The fact that women generally under

estimate their ages causes a biassed error in the average age of the population, while the tendency on the part of people to return their ages at the nearest round number causes an unbiassed error and, on the whole, does not affect the average age of the population materially. Biassed errors seriously affect the accuracy of the figures and should, therefore, be avoided. To eliminate the effect of unbiassed errors a large number of observations should be taken. So, if the errors cancel each other even a considerable degree of inaccuracy may be allowed while editing the data, but if they are cumulative they shall have to be avoided.

Approximation.

Numerous digits confuse the mind. They may be expressed in round numbers, even though exact figures may be available. For example, the figure 264,571 may be expressed as 264,570 or 264,600 or 265,000, but not as 264,500 or 264,000, or the population of India for 1941 may be expressed as 389 millions instead of 388,998 thousands.

Here arises another type of error called **Possible error.** If a quantity is rounded off to the nearest 100, the upper limit of error is + 50 and the lower —50. The possible error is therefore expressed as ± 50. That is, possible error denotes the upper and the lower limits within which the actual error lies. For example, the figure 1449 or even 1450 can be expressed as 14 hundred if approximated to the nearest 100, and, similarly, the figure 1351 or even 1350 can be expressed as 14 hundred if rounded off to the nearest 100. If, therefore, we are given only the round figure 14 hundred and not the actual figure of which it is an approximation, we may express 14 hundred as 1400±50, meaning thereby that the actual value of the estimate lies between 1450 and 1350, or that its possible error lies between + 50 and —50.

In approximation all figures except the first digit beyond the margin of accuracy should be left out. For instance, if the length of a leaf is recorded as 3.97 centimeters though with the scale used it was possible to read correctly to the nearest tenth of a centimeter it is better to retain the final digit 7 since 3.97 is a closer approximation to the real length than 4.0 cms which would be the reading if the digit 7 is dropped.

Again *if the last correct figure is a cipher it must be so entered.* If a leaf measures very nearly three centimeters it should be expressed as 3.0 cms and not simply as 3 cms for the latter figure might mean that the reading is accurate to centimeters whereas the former indicates that the reading is correct to millimeters which really is the case in our example. If the reading is correct to centimeters any leaf between 2.5 and 3.5 cms in length would be entered as 3 cms while if the reading is correct to millimeters the entry 3 cms would mean that the length is between 2.95 and 3.05 cms. Similarly an entry of 3.00 will mean that reading is accurate to hundredths of a centimeter and that the length of the leaf is between 2.995 and 3.005 cms.

When certain digits of a number are to be dropped *all fractions over half should be counted as whole numbers and all under half discarded.* Fractions equalling exactly one half may be allowed to remain or left out at discretion. The following table giving approximations should be carefully studied.

Original Number	Approximation correct to one decimal place
15.049	15.0
15.050	15.1
15.249	15.2
15.257	15.3
15.948	15.9
15.951	16.0

Editing Secondary Data.

Secondary data should be used only with careful inquiry and criticism. It is never safe to take them at their face value without ascertaining their meaning and limitations. Inquiries relating to the following should be made before they are used:—

(1) The organisaton that supplies the data.

(2) The reliability of the compiler, and his ability to procure correct figures.

If upon these two inquiries it is found that the data are not mere guess work, but are sound, the following information should be had:

(3) The scope and object of the inquiry conducted by the original compiler.

(4) The definition of units in which the data are expressed.

(5) The sources of the compiler's information.

(6) The method of his collection, including instructions given to the enumerators.

(7) The degree of accuracy desired and achieved.

(8) The extent to which they refer to homogeneous conditions.

(9) The suitability of their application to the given problem.

Even when all the inquiries have been thoroughly made, there may still be some shortcomngs in the data. The investigator using the secondary data will then have to exercise his common-sense and experience in using them. If the organisation or compiler's standing were not well-known, the investigator could study such part of the compiler's data as was familiar to him, and detect if there were any bias or motive to manipulate the figures. A business intelligence bulletin,

for instance may be very often biassed Data may be lacking in consistency or homogeneity or may not be suitable to the inquiry in the investigator's hand At every step then investigator s intelligence counts much Indeed as Dr Bowley remarks In collection and tabulation commonsense is the chief requisite and experience the chief teacher Where compulsion was exercised by the government the informants might have given information reluctantly and no wonder not precisely Questions relating to use of intoxicants mental infirmity total earnings profits etc. are inadequately answered since they arouse antagonism and suspicion among the informants The agency for collection may have been inefficient methods defective accuracy wanting The data may not have been co-ordinated They may have been collected for administrative purposes where a high degree of accuracy was not essential A sifting inquiry is, therefore necessary If co-efficients were computed it should be seen whether the numerators and denominators were related to each other and were homogeneous It should be known whether the quantities were strictly measurable and whether they were measured on the same basis In editing secondary data for the purpose of one s inquiry the investigator must be cautious and careful at every step The best thing to do will be to see whether the details compiled by the other organisations or compiler tally or tolerably agree with the details that would have been followed by the investigator himself if he personally conducted the inquiry

EXERCISES

(1) What precautions should be taken in approximating large figures?

Approximate the following figures expressing them (i) correct to nearest 1000 and (ii) correct to next 1000 over and show the biassed and unbiassed errors in approximation —

485,399; 410,902; 415,500; 290,492; 365,432; 399,491; 450,256; 462,300; 300,099; 295,591.

(2) Give examples to show that

(*a*) biassed errors are cumulative and unbiassed ones are compensating;

(*b*) relative accuracy is more important to a statistician than absolute accuracy;

(*c*) it is conventional accuracy of measurement that is generally looked for.

(3) In what way does a statistical error differ from a mistake? What classes of errors are there, and how may they be measured?

(B.Com., Alld., 1943).

(4) What precautions should be taken in making use of published statistics for further investigation.

(B.Com., Agra, 1939).

(5) (*a*) Discuss the main sources of errors in statistics and their effects.

(*b*) State the important methods of approximation and their utility in statistics.

(B.Com., Agra, 1940).

(6) 'Let us have quantity as well as quality in statistical data; but if there be a choice between them, the latter is more important and essential than the former'—Explain.

Is the above a good maxim for one editing the collected data?

(7) What do you understand by editing of data? What will you do if the data before you are

(*i*) incomplete but representative,

(*ii*) incomplete and unrepresentative,

(*iii*) certainly wrong,

(*iv*) probably wrong?

(8) It is never safe to take published statistics at their face value without knowing their meaning and limitations —Bowley

Elucidate the above statement and point out the general rules that you would lay down for making use of published data

(9) Of the biassed errors the statistician should have none, but of the unbiassed ones the more the merrier notwithstanding that they are also errors Elucidate

(B Com All 1947)

CHAPTER VII

STATISTICAL MATERIAL IN INDIA

Chief Sources of Economic Statistics.

The chief sources of secondary economic statistics available in India are the periodic reports and publications of (1) Central and Provincial Governments, Indian States, District and Municipal Boards, (2) Committees and Commissions appointed by the Government, (3) Semi-government institutions like the Railways or the Universities, (4) Research agencies, (5) Trade associations and private organizations, (6) Technical periodicals. A list of some of the important official and non-official publications is given in Appendix II. Non-official publications do not necessarily deal with non-official statistics. Rather, they mostly give official data. Non-official statistics obtaining in the country are meagre. They consist of statistical compilations of certain trade associations and other institutions which, in many cases, are regarded as highly confidential and, in several others, do not see the light of the day. They also consist of a few village and marketing-surveys and other statistical inquiries organized by some universities and other bodies in India in recent years, and by a handful of individuals in the past. So many of them remain unpublished. All these studies, no doubt, indicate the lines along which much useful work can be done, but the private individuals and institutions do not generally possess such facilities for collecting reliable information as the government departments do, and the latter alone are usually able to meet the cost of publication. Therefore, the major bulk of the statistical material available in India comprizes of official statistics, the collection of some of which—particularly of

those relating to prices, land values and cultivation costs—dates back to the early nineteenth century Great care and caution must, undoubtedly, be exercised in using non-official statistics but even the official statistics are not suitable for scientific purposes without a searching enquiry, since they are the by products of administrative machine And, the conflict between statistical needs and administrative purposes is very well known

Short-comings of Official Statistics.

Indian Official Statistics are of **questionable accuracy** for two main reasons In the first place, the agency for the collection of primary data is hardly trustworthy For instance, collection of agricultural statistics, prices and wages is done by the least qualified men in the Revenue and Police Departments of the Provincial Governments In the second place, scientific methods are very rarely applied to the analysis of primary data

Economic statistics in India suffer from **inadequacy.** The Indian Economic Enquiry Committee 1925 examined the material then available for estimating the economic condition of various classes of people in India and concluded as follows

For the purpose of determining in what respects the statistical data available are deficient from economic point of view the subject may be considered under the following three main classes

(1) General statistics other than production comprising Finance, Population, Trade, Transport and Communications, Education, Vital Statistics and Migration
(2) Statistics of Production including Agriculture, Pasture and Dairy Farming Forest, Fisheries, Minerals, Large Scale Industries, Cottage and Small Scale industries
(3) Estimates of Income, Wealth etc. Income, Wealth, Cost of Living Indebtedness, Wages and Prices

The statistics falling under Class (1) are more or less complete, those under (2) are satisfactory in some respects but

incomplete or totally wanting in others while as regards estimates of income, wealth, etc., under Class (3) no satisfactory attempt has been made in India to collect the necessary material on a comprehensive scale."

The language of the Bowley-Robertson Committee with regard to the nature of statistical data available in India cannot be improved. They wrote in 1934:

"The statistics of India have largely originated as a by-product of administrative activities, such as the collection of land revenue, or from the need of information relating to emergencies, such as famines. Only in the case of the population census and to some extent of foreign trade has there been an organisation whose primary duty is the collection of information. As a result the statistics are unco-ordinated and issued in various forms by separate departments. Though in some branches careful work is being done and determined efforts made to improve the accuracy and scope of information, in others they are unnecessarily diffused, gravely inexact, incomplete or misleading while in important fields general information is almost completely absent. The only co-ordinated general publication is the Statistical Abstract, which omits some important statistics which must be searched for in other documents. The situation cries out for overhaul under the control of a well-qualified statistician".

Indeed, from the multifarious activities of state in India there springs a constant and copious stream of numerical data at regular intervals. **Inconsistency** and **incompleteness** in existing statistics are natural in a country where the task of collection and presentation of official statistics devolves on the different departments of the Government of India, Provincial Governments and Indian States with their own administrative needs and personnel.

Lack of co-ordination is another glaring defect of Indian economic statistics. Co-ordination of official statistics is not possible in the absence of a co-ordinating authority existing in the country. Most foreign countries possess Central Statistical Bureaux for collecting and editing all statistical matter of public

interest The establishment of such a bureaux was recommended by the Indian Economic Enquiry Committee and that of a Permanent Economic Staff by the Bowley Robertson Committee Neither of these recommendations has yet been fully carried out The present statistical organisation of the Government of India consists of a Director of Commercial Intelligence and a Director of Statistics at Calcutta responsible for collection of some official statistics and an Economic Adviser to the Government of India in New Delhi to undertake collection and interpretation of some other statistics There is also a Statistician to the Government of India while some Government departments have their separate officers in charge of statistics There is no department of statistics as such at the headquarters of the Central Government The need for the establishment of a Permanent Economic Staff by the Government of India to collect and co-ordinate various statistics in the country was never more urgent than now, when they are committed to chalk out plans for the economic development of India in the post war period and in pursuance of which they have established a Department of Planning and Development

Yet another short-coming of official statistics is that the exact significance scope and method of their compilation are not widely known so that they are not self explanatory Light on these points however in regard to statistics of production and prices is now thrown by the *Guide to Current Official Statistics*, Vol I

Delay in publication of even the inadequate and defective data simply adds insult to injury Figures become completely out of date by the time they are published and thus much of their usefulness is lost

The Bowley Robertson Committee made some valuable suggestions for improving the official statistics of the country These recommendations were considered for long by the Government of India The scheme of an economic census and the

establishment of an Economic Staff were regarded as too costly and abandoned for the time being.

Examination of some Official Statistics.

It would now be well to examine in detail some of the important official statistics with a view to study their short-comings and offer suggestions for remedying them.

Statistical Abstract of India.—In this publication all important economic statistics, including, among others, those concerning finance, currency, banking, population, industry, communications, labour and insurance, relating to India, and where available, to Indian States, are regularly published. These data are based on the information furnished by different departments of the Central and Provincial Governments and Indian States. The *Abstract*, therefore, contains all those errors from which the original compilations suffer. The short-comings of this document are:

(1) Lack of adequate co-ordination, though some co-ordination is being attempted.
(2) Lack of completeness and consistency in existing figures.
(3) Delay in publication.

It is suggested that a suitable machinery for proper co-ordination such as that proposed by the Bowley-Robertson Committee should be set up. The Committee have recommended the establishment of a Permanent Economic Staff consisting of two trained economists, one statistician and a Director of Statistics. The Director's duties include co-ordination of central and provincial statistics. For the second short-coming, the Committee have made valuable suggestions, which only need implementing. Information about some important problems which so far is not available in the *Abstract* should be included. Delay in publication is partly due to the inclusion of less

important items which hold up the publication. If unimportant items are deleted the publication would not only become handy and less expensive but would also be published without unnecessary delay. It has been argued that if the *Abstract* is divided into different parts and each part is published as soon as information relating to any section is available the defect of delay shall be greatly reduced. But in view of the separate publications dealing with different subjects such as *Statistical Tables relating to Banks* and *to Progress of Co operative Movement in India* already existing this argument may be met if the publication of these special reports is speeded up. It may however be suggested that like the procedure adopted in the *Year Book of the League of Nations* estimates in place of finally revised figures where the latter involve undue delay may be published the fact being mentioned in a footnote. Thus delay in the publication of the *Abstract* would be very much reduced, and its usefulness increased.

Agricultural Statistics.—These statistics deal largely with the land utilized for agricultural purposes and the crop raised on it. Since land revenue has been the one important source of revenue in India we possess valuable statistical records relating to land crops and yields since so early times as the famous settlement of Todar Mall, the Revenue Minister of Akbar. The late administrators of India collected agricultural statistics at an early stage particularly when they introduced the *Ryotwari* system towards the close of the 18th century. Provincial Governments took up the compilation of agricultural statistics in 1866. In 1885 first crop forecast relating to wheat was made followed in later years by forecasts of cotton oilseeds rice jute groundnuts and sugarcane.

Agricultural statistics are published regularly by the Department of Commercial Intelligence and Statistics in the following annual publications

(1) *Agricultural Statistics of India Vols I & II*
(2) *Estimates of Area & Yield of Principal Crops in India*

(3) *Plantation* (Tea, Coffee and Rubber) *Statistics.*
(4) *Summary Tables of Agricultural Statistics.*
(5) *Statistical Abstract for India.*

In addition to the above, *Crop Forecasts* and *Intermediate Crop Forecasts* are periodically issued, while *Report on the Census of Livestock, Ploughs and Carts in India* is a quinquennial publication. Besides, agricultural statistics are also available in the *Indian Trade Journal* (weekly), *Monthly Survey of Business Conditions, and Quinquennial Report on the Average Yield per Acre of Principal Crops in India.*

The difficult task of making primary estimates relating to agricultural statistics in India falls on the officers of the Provincial Revenue Departments, who have to carry it out amidst their heavy administrative and revenue-collecting duties. It needs no emphasis that they have neither the time nor the necessary training for the work entrusted to them. Consequently, the reliability of Indian crop statistics is of doubtful character.

In areas where *Ryotwari* system or Temporary Settlement prevails all villages have been carefully surveyed and mapped. There, the village accountant, called *Karnam* or *Patwari* keeps field records. At the beginning of the sowing period he prepares a statement showing areas under different crops in his village, and submits it to the Revenue Inspector. These figures are aggregated in *tehsils* or *taluks,* districts and provinces. Once or more during the growing period, and finally at harvesting, the village accountant estimates the yield of the crops as so many annas, generally taking 12 to 16 annas as standard. The *Tehsildar,* exercising his general knowledge of the condition of crops, reports a single result for all the villages under his jurisdiction to the District Officer. The latter modifies these figures in the light of his knowledge or discretion and reports a single number of annas, or an average, for the district. This

average is very vaguely defined and leads to suspicion Usually it is the mode Further, it has been found that in many cases the local official does not visit the fields but deduces the area from the quantity of seed the cultivator says he has sown This malady is further aggravated by the indifference of the revenue officers towards checking the *Patwari's* figures personally With regard to the *annavari* estimate of the yield, it may be pointed out that the estimate is vitiated in some cases by the failure of the revenue officers to actually get crops of a small area from an average field cut and compared with the standard yield Again since remission of land revenue has to be granted in temporarily settled areas where the seasonal condition falls below a certain percentage of normal it is not impossible that the village accountant and subordinate officers may pitch their *annavari* estimate too high when the seasonal condition is on the border line for the grant of remissions The village accountant is said to possess a bias to report no change from the previous year to under-estimate a good crop or to exaggerate the fall in a bad crop The margin of error in his estimates is an unknown quantity Efficient supervision due criticisms and proper scrutiny of the *Patwari's* estimate of both area and yield are essential if reliable and useful data are to be collected.

The Department of Agriculture in each province fixes the normal yield per acre for the different crops in each district on the results of crop cutting experiments conducted for each crop sown on plots of average quality These experiments are too few and the plots that are singled out for the experiments are selected according to purposive selection in which personal bias has a great chance of prejudicing the choice The normal yield therefore loses its representative character The *annavari* estimates of a particular year are compared with such defective normal yields The comparison is vitiated According to the Bowley Robertson Committee if the direct

method of estimating the yield in maunds per *bigha* or bushels per acre is adopted, dependence on the standard yield would be completely done away with and accuracy of data would improve. But, since the direct method is said to be impracticable, the yield should be stated to the nearest anna for each village and weighted arithmetic mean of these statements should be obtained for a *tehsil* or *taluk*. This mean would be fairly reliable. The condition factor for each *tehsil* or district could be expressed as a percentage of the normal rather than as so many annas. It is further suggested that for the computation of a reliable normal yield a large number of plots should be selected on 'random sample' basis. It is worth noting that improvements in this direction are taking place in the U. P. and a few other provinces.

In permanently settled areas, such as those in Bengal and Bihar, there are no village accountants and subordinate revenue officials, nor are the villages surveyed and mapped. There, the Revenue Officers have to depend for estimates of area and yield on the guesses of the village headman supplied to the former through police officers. The Revenue Officers are required to check these figures from personal observation during their tours. They have not the time to do it always. These guesses are mostly under-estimates. It is suggested that the system of printed forms, in vogue for collecting figures relating to area under jute, should be adopted for other crops in these tracts.

It may be pointed out that statistics of area under different crops are fairly trustworthy in temporarily settled areas, since on their accuracy depends the collection of land revenue. But similar figures for permanently settled areas are far from satisfactory, since they are not required for revenue purposes. The Bengal Famine of 1943 proved beyond doubt the untrustworthiness of the agricultural statistics of the province, and it is only proper that the Government of Bengal have lately decided to so improve their statistics as to bring them in line with

those of the temporarily settled areas Greater improvement however is called for not only in Bengal but throughout the country including Indian States

Agricultural statistics available at present in India, because of their short comings are quite insufficient to determine whether food is increasing in proportion to population It is not possible to deduce from them the quantity or value of total agricultural produce This presents a serious handicap for economists and statesmen to tackle food problem in the country Even the yield figures are not sufficiently reliable and areas for minor or mixed crops are not separately known.

The Department of Commercial Intelligence and Statistics which publishes crop forecasts two or three times for the different commodities mentioned above in their respective seasons bases them on the information primarily supplied by the village accountant If *Patwari's* bias can be removed and his work properly supervised the accuracy of his data and with it that of crop forecasts will improve Then the usefulness of forecasts to commercial community would certainly increase a good deal Besides crop forecasts for other commodities like *jowar, bajra* maize should also be made In order that reporting may improve the Agricultural Departments should be given an increasing share in making general estimates of yield.

A quinquennial census for livestock is taken in different provinces But the information relating to animal products—milk meat, eggs hides bones etc.—is very little A detailed knowledge is desirable Further the classification of cattle should be amplified to furnish greater details The fact that some of the provinces and Indian States keep away from the quinquennial census on one pretext or the other itself speaks of the indifference with which this vital problem is treated A uniform policy for the whole country is very necessary

Industrial Production Statistics.—The following publications give statistical information relating to large scale industries in India:—

(1) *Monthly Statistics of the Production of Certain Selected Industries in India*—(six in a year).
(2) *Monthly Statistics of Cotton Spinning and Weaving in Indian Mills.*
(3) *Large Industrial Establishment in India.*
(4) *Statistics of Factories*—(Annual) issued by the Labour Department, Government of India.

In addition to the above, *Indian Trade Journal, Statistical Abstract and Monthly Survey of Business Conditions* also afford data relating to industrial production.

Statistics available in the above publications include the number of factories, the average number of workers employed daily, the quantity of cotton yarn and cotton manufactured goods, jute, paper and iron and steel manufactures, production of petrol, kerosene oil, cement, sugar, matches, heavy chemicals, spirit, paints and wheat flour.

These statistics refer only to factories employing twenty or more persons per day. No attempt has been made to collect all-India figures of production of the cottage industries of the country. The only cottage industry favoured in this regard is the hand-loom industry for which estimates of production are published in the issues of the *Review of Trade of India.* Nor is any information available with regard to the number of workers engaged in cottage industries except in the not quite satisfactory occupational tables of the 1931 *Census Report*, which too were retrenched in 1941. Further, the returns from Indian States are not complete, and it is difficult to understand why the published statistics exclude the production of such important industries as soap manufacture and vegetable-products industry. The available data relating to industrial production are therefore meagre.

Consequently it is difficult to study in all its aspects any economic problem involving reference to the country's industrial output.

Prices, Wages and Cost of Living.—Market prices of staple commodities for smaller towns and wages for some grades of labourers are reported regularly in the *Gazette of India* and *Provincial Gazettes*. The revenue officials are required to report prices. Overburdened with their administrative duties they hardly have sufficient time to collect price quotation. Adequate attention is not paid to exact description of the grade and distinction between wholesale prices for small lots and retail prices is not made clear. Sometimes prices of the same commodity but of different qualities and at other times even when actual prices have changed the same old prices are quoted. In order that correct information may be had dealers in staple commodities should be persuaded to send regular quotations of the same commodity of the same quality. These quotations may be verified by proper supervision.

Wholesale prices of certain agricultural and non-agricultural articles for a few bigger towns of India are available in a *Monthly Statement* issued by the Department of Commercial Intelligence and Statistics. Prices for the past few months are also shown along with current prices so that comparison can be easily made. But when at times continuity in monthly quotations is broken the benefit of comparability is lost.

The General Index Number of All India Wholesale Prices published until recently had outlived its utility, since its base year was so old as 1873 the list of commodities had not been revised since 1889 and it was unweighted. Its publication has been discontinued since August 1941. Index Numbers of weekly wholesale prices of certain articles in India' with week ending 19th August 1939 as base are being regularly published now in the *Monthly Survey of Business Conditions in India*

issued by the office of the Economic Adviser, Government of India. These index numbers are based on only 23 commodities and are unweighted. Geometric mean is used in their computation. Against them stand the Calcutta and the Bombay Wholesale Price Index Numbers, which are based on much larger number of items and are weighted. Naturally, these index numbers, all representing wholesale prices in India, register a huge difference for the same month. The base of the Calcutta and the Bombay Wholesale Price Index Numbers is July, 1914, which should better be changed now. In addition to these index numbers, wholesale price index numbers for Madras and Kanpur are also available in the *Monthly Survey*.

Wage statistics can be classed into three categories—(1) Factories and Mines, (2) other Urban Occupations and (3) Rural Occupations. The task of collecting statistics for the first two categories should be entrusted to Provincial Labour Officers who may be appointed where they do not exist. Bombay's example is commendable in this regard. Regarding wages of urban occupations, scanty attention has been paid to those working outside the factories in towns, *e.g.*, municipal employees, artisans, porters, builders. The range of occupations included is not comprehensive for towns. Wages for rural occupations are often quoted between wide range and the frequency of employment is not indicated. Classification of urban and rural workers is inadequate. Therefore, an idea of general movement of rural wages is difficult to obtain. It may be suggested that in each district a small number of villages, where wages are paid in cash and separate occupations are few, should be selected. Wages paid for the different occupations should be collected, care being taken to see that in each successive record the wages are paid for the same work and are strictly comparable. The unweighted average of the rates for each occupation would afford a fair measurement of the general movement of rural wages in a province. Wage rates

should not be quoted as varying between two limits and frequency of employment should so far as possible be ascertained

Cost of Living Index Numbers are now available for 27 different towns of India They are published in the *Monthly Survey of Business Conditions in India* but not in the *Statistical Abstract* This omission should be rectified Besides cost of living index numbers should be computed for other labour areas where wage-payments are made on a cash basis so that public opinion may be kept well informed particularly when a wage dispute turns on the expenses of living Further, a separate index number for salaried persons, say under Rs 100 a month is worth attempting

Trade Statistics—Statistics of foreign trade are available in the monthly and annual *Accounts and Statement of the Sea Borne Trade and Navigation of India* as also in the annual *Review of the Trade of India* published by the Department of Commercial Intelligence and Statistics They generally give the information that is practicable regarding the exports and imports of Foreign Merchandise, exports of Indian Merchandise and total Exports under five main classes These classes are (1) Food Drink and Tobacco (2) Raw materials and produce and articles mainly unmanufactured (3) Articles wholly or mainly manufactured (4) Living animals (5) Postal articles But since imports and exports on Government Account are not given in as much detail as those on Private Account it is not possible to arrive at the total trade in particular commodities This short coming should be made up

Statistics relating to inland trade are now available in the *Accounts Relating to the Inland (Rail and River borne) Trade of India* issued by the Department of Commercial Intelligence and Statistics Similar statistics were published by the Department of Statistics upto 1922 The present series retains in essentials the form of the older publication and is a monthly production

The trade dealt with in the publication falls into one or other of the following categories:—

(i) Trade of a province with other provinces,
(ii) Trade of a chief port or other ports with the province in which such port or ports are situated, and
(iii) The trade of a chief port or other ports with other provinces and ports.

In addition to the above publication, *Accounts relating to the Coasting Trade and Navigation of India, Trade Statistics relating to the Maritime States in Kathiawar and the State of Travancore* are also monthly published by the Commercial Intelligence and Statistics Department. Thus a good account of India's trade statistics is available.

The Census Reports.—Census of population is generally taken every ten years in India. Sometime before the date fixed for the census, a Census Commissioner is appointed, who selects Superintendents of Census for each province and native states. The Superintendents select honorary Census Supervisors and Enumerators for each district or locality. Enumeration is done through the municipality in a town and through the *Tehsildar* in a rural area. Indian census is unpaid. To take the preliminary census the enumerator visits every house in his block sometime before the census day, and collects the required information from the head of each family. Then, on the night fixed for the census there is a simultaneous country-wide count. The collected data are scruitinized, classified and tabulated to produce the census reports. These reports give valuable information relating to the distribution and density of population, vital statistics, urban and rural population, age and sex distribution, civil condition, infirmities, occupation, literacy, languages, etc.

The last census was taken on 1st March 1941 It puts the estimate of population at 389 0 millions [1] of which 87% is rural and 13% urban The one-night enumeration which was adopted upto the census of 1931 was in 1941 replaced by a period system It enabled the number of enumerators to be halved as against the number employed in 1931 Schedules used formerly were abandoned Instead enumeration was carried out directly on the slips which were later sorted to produce tables A new feature was the taking of 1/50 random samples of the entire population to be used for making several deductions Religion as criterion of *census differentiation had several drawbacks and was substituted* by the concept of community And, some such new questions as the age of mothers at the birth of their first child and the number of children born were introduced Returns for this question would help the computation of net reproduction rate for the country

The short-comings of the census reports are many There is a marked lack of uniformity in the classification of occupations from census to census A study of occupational structure of the country is therefore, rendered difficult Further the distribution of industrial workers into employees and those working on their own account is not available Nor are the parts of the year for which a worker is employed and the parts for which he is unemployed known Under the circumstances it is difficult to formulate a sensible occupational plan for the country

Indian age-returns are admittedly inaccurate mainly because of the ignorance on the part of the informants of their precise age Besides ignorance, there are some psychological reasons too The age-period of girls from 10 years to 15 years is defective in numbers because of the unwillingness of some people to admit having unmarried daughters, who, according

[1] Inclusive of Burma the total population would be over 400 millions.

to custom in the community or religious injunction, should have been married by then. For widowers and bachelors, particularly if they have a wish to remarry, there is a tendency to under-estimate their ages. Recently married girls and particularly those who have become mothers tend to over-estimate their ages. The old people have an inclination to overstate their ages. Again, there is a marked preference for stating the age at a digit ending with zero or five. Enumerators are instructed to correct ridiculous returns of age. If they are conscientious, they do it by asking the person concerned questions about his age at the time when some well known event in the past occurred. But with the limited ability of the enumerators it is highly doubtful that they are in all cases competent to detect the wrong and verify it. Indian custom permits only the female investigator to verify the information about *pardanashin* ladies. Such investigators should be appointed.

Returns for civil condition in 1931 exhibited an excess of married males over married females, whereas in the previous censuses the ratio was reverse. The reason for this was the promulgation of the Child Marriage Restraint Act (Sarda Act) in 1930, because of which those people who had married their male children under 18 years and female children under 14 years may have hesitated to disclose the truth for fear of prosecution.

Infirmity is very much concealed, particularly among females. Deafness of children is, in many cases, kept a secret. Insanity and blindness are generally matters of personal temperament. Leprosy is infectious at an early stage when it cannot be detected by the lay eye of the enumerator.

Accuracy of census returns depends to a large extent on the capability of the enumerator and also on the general circumstances prevailing in the country at the time the census is taken. The census enumerator constitutes the front-line force, and if he is not given adequate training he cannot be

expected to hit the mark with precision in all cases It is no wonder if the blank forms used in 1941 could not be filled in correctly by many an enumerator since the forms were not simple and the enumerators were not given adequate training Training of the enumerators is no doubt necessary but along with it supervisors too must also be picked up from among those possessing a knowledge of statistical methods and ways of conducting demographic enquiries

Circumstances obtaining in the country at the time of census operations have a great part to play The effect of Sarda Act has already been pointed out. Further when the distribution of seats in the legislatures local bodies and government services is based on the relative strength of the persons belonging to different religions it is not improbable that some people exaggerate the number of members in their family—something which an inefficient enumerator cannot always detect This factor is believed by some people to have caused an over estimation of the population in 1941 Again if secrecy of census returns is not ensured and people have the apprehension that the returns may be produced in a court of law as evidence of age or of a marriage against the provisions of the Sarda Act, wrong returns would naturally result Secrecy should be ensured. The Census Act should be made permanent and a permanent staff should be appointed at the centre to work up the material collected at each census The present system of setting up the whole census machinery hastily before the census and disbanding it after the publication is complete whereby the experience gained at one census is not fully utilized at the other should go Essential information for each district should be made available in booklets Tabulation should be done by machines.

The above suggestions are not meant to be exhaustive but they do indicate the lines along which improvements can

be effected in the Indian census. Improvements are necessary because the census has great potentialities. Of course, the census is primarily meant for administrative purposes, but it also affords much valuable information for the economist, the sociologist and the businessman. The economist, for instance, can study, on the basis of census figures, the population trend of the country, its occupational structure, and the increment in urban population. Utilizing other relevant data along with these studies he can trace the correlation between population-growth and food-supply, between occupational changes and the effect of granting protection to industries, and between increment in urban population and decay of rural crafts. The sociologist may study the possibilities of effecting reforms in respect of, say, ages at which people should marry, or arrangements that should be made to bring down infantile mortality.

Businessmen do not always realize that the census reports contain information which is of an important nature for them. If they do, many of the problems they are confronted with can be properly attended to. India has a very large volume of internal trade, and every human being returned in the census is a *consumer*. To the businessman a knowledge of his consumers and their location is evidently of immense aid. Again, with a knowledge of the density of population of different areas an estimate of the areas where development of markets is likely can be made. The higher the density of a certain locality, the greater is the market there. The selling cost will always be low since delivery service in a dense population is cheap. Further, knowing the number of inhabitants in a town and the quantity of goods the businessman had been usually selling there, it would be possible for him to compute the *per capita* consumption. And, if this *per capita* consumption shows a fall for no valid reasons other than lack of efficient push the businessman can launch an intensive selling campaign to increase his sales. The

class for wh ch his goods are specially meant say ladies infants mil tary people can be approached in a business like manner Besides, occupational statistics would tell the businessman whether a certain area is inhabited by the poor whose purchasing power is low He should then see if his goods are meant for such a class If not he would be wasting money over trying to gain a foothold in that particular area He would also be able to gauge the present and future supply of labour on the basis of occupational statistics and make adjustments accordingly

A transport agency say a railway would find valuable nformation in the census reports The area which is densely popu lated or would be so populated if only the means of transport are improved or introduced should receive the first attention of the transport authority Such areas would also be good for advertis ng agencies to push their advertisement among the cl ss of people to which their goods relate Producers of staple commodities and manufacturers of industrial goods can equally benefit from the material collected at the time of census operations If the population of a town falls demand in that area will fall unless the demand of the existing population rises proportionately Chang s in the sex ratio in occupational structure or in age ompos tion are likely to affect the demand for goods Similarly l fe insurance companies may compare their estimates of expec tation of life with those published in the census reports and see that their premium rates are properly drawn up The legisla tors shall be able to study the necessity of framing leg slative provis ons for removing the ills from which society suffers upon a study of infant mortality fertility rate sex ratio infirmities Numerous other uses of census figures can be thought of All that has been said so far ind cates the immense value of the census reports to d fferent people It is imperative therefore to make the Indian census reports up-to-date and as useful as possible

Vital Statistics.—The vital statistics of India are admittedly defective Figures published in the *Statistical Abstract* are

definitely misleading. The system of registration of births and deaths varies in different provinces. Generally, they are kept up by the reports of village officials in rural tracts and by municipalities in the urban areas. The reporting of births and deaths is an irksome duty which the village headman, illiterate and ill-paid as he is, often neglects. In the case of births he is very likely to wait and see whether the child would remain alive to save himself of the worry of making a second report of its death if death soon occurs. He hesitates to report deaths to avoid the unwelcome visits of the unduly suspicious police officers. He generally holds up the reports of births and deaths for a weekly or fortnightly visit to the *tehsil* or *taluk* headquarters. The records in towns are said to be more imperfect than those in villages. A clearer appreciation of the population problem of India would have been possible, if only vital statistics were accurate and complete. The census reports of 1941 might be able to throw some light on the net-reproduction rate in India, without which knowledge it is impossible to lay down any sensible population policy for the country. The system of registration of births and deaths should be brought up to the level reached in other countries. It may be suggested that the list of persons to whom such daily occurrences may be reported should be widened, and an organisation should be established at the head-quarters of the Central Government to deal with these day to day instances.

National Income.—There are no official estimates of the national income of the country. The unsatisfactory character of the statistical material available in India for measuring the national income, as found by the Indian Economic Enquiry Committee in 1925 and the Bowley-Robertson Committee in 1934, has already been referred to, and some of the major deficiencies that exist even to-day have been pointed out in the foregoing pages. It is unfortunate that the valuable suggestions and plans set out by these two committees for measuring the

national income of the country have all foundered on the rock of financial stringency

National income is the money measure of the *net* aggregate of commodities and services accruing to the inhabitants of a country during a year Three methods of measuring it are available Census of Production Census of Incomes and Occupational Census Of these not even one method can be followed to its full in India so long as the statistical data remain as meagre as they at present are and also so long as no central statistical organisation is set up to supervise and co-ordinate economic surveys intensive and extensive rural and urban The suggestions made by the Bowley Robertson Committee for the measurement of national income need now when economic planning is the Government's declared policy more than ever before a close and sympathetic attention of the Government A summary of these suggestions is given in Appendix III

Inspite of the admittedly insufficient and not very reliable statistical information estimates of national income and income *per capita* for India have been attempted by various people from time to time In using them it must not be forgotten that they are mere estimates and are based on deficient statistics The following table should be carefully studied in this connection

Estimates of Income per capita for India

Estimated by	Year for which estimated	Estimated income *per capita* in Rs
Dadabhai Naoroji	1867 70	20
Earl Cromer & Sir David Barbour	1882	27
William Digby	1898-99	17 5
Lord Curzon	1900	30
Wadia & Joshi	1913 14	44 5
Shah & Khambata	1921 22	67
Findley Shirras	1922	116
V K R V Rao	1931 32	65

EXERCISES

(1) Explain, with examples, the important sources of errors in the census returns. How can these errors be avoided?

(B.Com., Alld., 1933).

(2) In what respects are the statistical data, available in India, deficient from an economic point of view? How can this deficiency be removed?

(3) In the Census Report for 1931, the Census Commissioner for India observes:—

'The error in the numerical count has been put at a maximum of one per *mille* and is probably less.'

Comment upon this statement.

(B.Com., Alld., 1935).

(4) What are the present sources of wage statistics in India? What measures may be adopted for removing their gaps and other defects?

(M.A., Patna, 1940).

(5) Explain the main defects of the statistics of prices and wages in India. How can these defects be removed?

(B.Com., Alld., 1938).

(6) 'The statistical publications relating to the decennial censuses of population in India leave little to be desired.'

(Indian Economic Enquiry Committee Report).

Comment.

(7) Discuss the method recommended by the Bowley-Robertson Committee for the measurement of the National Income of India.

(B.Com., Alld., 1940).

(8) What methods are usually adopted for estimating the national dividend of a country? To what extent, in your opinion are recent estimates of the 'national dividend' of India reliable?

(M.A., Alld., 1935).

(9) Why is an economic survey of a country considered essential before adopting a programme of economic development? How would you conduct an economic survey of India?

(M A, Alld, 1935)

(10) Describe briefly the nature and sources of the data used in the Review of the Trade of India. Are there any gaps and defects in this account?

(M Com Luck, 1942)

(11) Explain the important sources of biassed errors in the collection of data regarding wages prices and yield of crops in India. How can these errors be avoided?

(B Com Luck 1938)

(12) How would you organise a Central Statistical Bureau for India? Explain clearly its main functions

(B Com, Luck 1938)

(13) Give the Blank form of a census schedule used in *India*. What improvements would you suggest to the schedule?

(B Com Luck 1939)

(14) Discuss briefly the methods of calculating National Income. How far are these methods available for calculating the national income of this country?

(B Com, Bombay, 1936)

(15) What statistical information is available in India with regard to (*a*) Imports and Exports (*b*) Prices and (*c*) Agricultural statistics? Examine their sufficiency

(B Com, Alld, 1943)

(16) Discuss the possible value of Census Reports to producers manufacturers and businessmen. How can the Indian Census Reports be made more useful to these people?

(M Com, Alld, 1943)

(17) What are the principal sources of statistical data for India? Examine suggesting improvements the materials available under the following heads (a) Agriculture (b) Wages and (c) Prices

(M Com Alld 1943)

(18) How will you estimate the wealth of a country? Discuss the problem of organizing a census of economic production in India.

(B.Com., Agra, 1939).

(19) Write a note on the inadequacy of statistical data in India for sociological and economic inquiries, suggesting methods for removing the inadequacy.

(20) State what you can about Indian Vital Statistics. Do they throw any light on the causes of India's poverty?

Or, Discuss the available sources of information in respect of India's trade, both foreign and inland.

(B.Com., Agra, 1940).

(21) What are the present methods of collecting agricultural statistics of acreage and yield per acre? Discuss the accuracy of the methods followed. Can you suggest improvements specially for permanently settled areas?

(M.A., Cal., 1936).

(22) Describe the present method of occupational classification followed in Indian censuses. Do you consider it satisfactory?

(M.A., Cal., 1936).

(23) Point out the main sources of error in the Indian Census Report, and suggest methods for avoiding these errors in future.

(M.A., Agra, 1943).

(24) If you are asked to compare the economic effect of the present war on India with that in Great Britain, what statistics would you use?

(M.Com., Luck., 1942).

(25) Discuss the utility of the data regarding occupations collected at the time of the last census. How can these data be utilised in estimating the National Income of India?

(M.Com., Luck., 1942).

(26) (a) Briefly enumerate the causes of high infantile mortality in India and suggest the steps to be taken to bring down the rate

(b) Examine the reliability and sufficiency of statistics relating to such infantile mortality in India

(B Com, Alld 1942)

(27) What do you understand by the net reproduction rate? How may this be utilized for estimating the future population of a country? Are there any special difficulties in the case of India?

(M.A Alld, 1942)

(28) What kind of information on social and economic subjects is available in

(a) The Monthly Survey of Business Conditions in India (b) The Statistical Abstract of India, (c) The Review of Trade (d) The Bombay Labour Gazette (e) The Indian Census Reports and (f) the Capital.

(29) Give the general method of preparing crop forecasts issued by the Department of Commercial Intelligence and Statistics in India Suggest measures for improving their accuracy and usefulness

(30) Examine the deficiencies in the following publications and suggest improvements to make them more useful to economists and businessmen in India —

(a) Statistical Abstract of India
(b) Area and Yield of Principal Crops in India

(M Com. Alld, 1944)

(31) The statistics even of crop production leave much to be desired while statistical informations about other important parts of agricultural income such as the output of animal husbandry are almost completely lacking and statistics of industrial production are patchy in the extreme (B R Committee Report)

Prove the correctness of the above statement by taking examples from Indian statistics and suggest measures for removing the defects

(32) Give in broad outlines the salient features of the Bowley-Robertson report on census of production in India.

(B.Com., Luck., 1944).

(33) Write a short essay on (*a*) Official Statistics in India, or (*b*) War and Statistics.

(B.Com., Agra, 1943).

(34 What are the sources from which statistical data relating to large scale industries in India can be obtained?

(B.Com., Nagpur, 1944).

(35) In what respects was the Census of India taken in 1941 different from that taken in 1931?

How far, do you think, are the returns relating to age, civil condition and infirmity made during census operations in India accurate?

(B.Com., Alld., 1944).

CHAPTER VIII

CLASSIFICATION AND TABULATION OF DATA

CLASSIFICATION

Statistical data, collected in the course of an inquiry concern a number of units of one kind or another which together form the group relating to the inquiry A statistical group consists of a large number of things or individuals having something in common but differing from one another in respect of some measurable characteristics For example students belonging to the same college may differ from one another in regard to their age civil condition or height But together they constitute a group A statistical group is very large so that no one can appreciate at a glance or even after a careful study the information relating to many units A reading of a thousand or more schedules returned by the students of a college respecting their age weight height, etc cannot enable the reader to get a proper idea of the details mentioned Some process of condensation must be devised for the purpose This process yields statistical tables But before tables can be prepared the different units must be grouped together into classes so that the like will go with the like and the unlike with the unlike Details would necessarily be lost since the individual units would be merged in a class For instance all those students who return themselves as married shall be placed in one class the unmarried in another and the widowed in the third From the table that shall then be prepared none shall be able to identify himself or herself since he or she shall be merely one in a class composed of those similar to him or her in respect of civil condition This process is called Classification

"Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities, and gives expression to the unity of attributes that may subsist amongst a diversity of individuals".[1]

The objects of classification are many. It clearly shows points of similarity and dissimilarity. It helps one to form a mental picture of the objects he can see or conceive. By condensing the details it saves one from mental strain. It affords an appreciation of the information that would otherwise have been left out as perplexing or unimportant. It prepares the ground for enabling comparisons and inferences. It institutes a logical and orderly arrangement of things.

Importance of classification in statistics cannot be overemphasized, and yet it is something for which no very precise rules can be laid down. Skill and patience are, no doubt, indispensable ; but as in collection of data so in their classification, experience alone will convince one of the requisite care that should be taken to avoid blunders and save time. It may be noted that an *ideal* classification should possess the merits of being unambiguous, stable and flexible. It should not leave room for doubt; it should be stable enough to render comparisons easy; and, it should be so flexible as to incorporate new ideas as they materialize in future.

Classification is determined by the characteristics possessed by the individual units of a group. These characteristics are of two kinds; descriptive and numerical. Descriptive characteristics comprize of attributes or qualities possessed by objects or individuals, such qualities not being quantitatively measurable. Characteristics like sex, civil condition, caste, religion and infirmity are descriptive. Numerical characteris-

---

[1] L. R. Connor, *Statistics in Theory and Practice* 1936 ed. p. 18.

tics are so called because they are susceptible of quantitative measurement Age height income weight are numerical characteristics Classification of given data by descriptive characteristics is generally called classification according to attributes while that by numerical characteristics is commonly known as classification according to class intervals

Classification according to Attributes

Descriptive characteristics can be classified by means of some natural or physical lines of demarcation Natural or physical differences determine the classes into which units should be placed It is easy in these cases to separate the similar from the dissimilar characteristics For instance population of India may be classed into male and female literate and illiterate blind and not blind Thus when one attribute is noticed two distinct classes are formed These two classes are exclusive of each other If the members of one class possess the common quality of being males those of the other are devoid of it A classification of this type where each class is divided into two sub-classes only is called **Simple Classification or Classification by Dichotomy**

Where more than one attribute is studied several classes may result For instance the population of India may not only be classed into males and females but males and females may be further sub divided into literate and illiterate such as male literate and female literate male illiterate and female illiterate Classification may be carried still further for instance according to occupation A male literate may be a teacher a female literate a stenographer a male illiterate may be a peon; a female illiterate a maid servant Further classification according to religion or caste is yet possible Numerous classes may thus be formed A classification of this type where each class is divided into more than two sub-classes is called **Manifold Classification.**

The following brief classification of languages of India affords a good example of manifold classification:

A. *Languages of India and Burma.*
(i) Austric, (ii) Tibeto-Chinese, (iii) Dravidian (iv) Indo-European, (v) Unclassed.

B. *Languages of other Asiatic Countries and Africa.*
(i) Indo-European, (ii) Tibeto-Chinese, (iii) Semitic, (iv) Hamitic, (v) others.

C. *Languages of Europe.*
(i) Indo-European, (ii) others.

It is necessary to state that in classification according to attributes the boundary line between different classes, though artificially set, is definitely made before the work of classification begins. For instance, the decision as to who would be recorded as literate and who as illiterate is made before actual classification.

Classification according to Class-Intervals.

Numerical characteristics can also be classified by assigning arbitrary limits. The ages of persons, for instance, are of indefinite variety; so also are heights or weights. But, the entire range of ages, heights or weights, from the lowest to the highest, can be broken up by drawing arbitrary boundary lines, and those units which are nearly alike in respect of a particular character put together in one class. Thus if the ages of a given group of people vary from 25 to 44 years and it is desired to divide the group into four classes, the boundary lines would preferably be fixed on numbers 25, 30, 35, 40 and 45 years. These boundary lines are known as the **class-limits,** the group constituted by two limits as the **class-interval,** the distance between two limits of a class-interval as its **magnitude,** and the number of observations falling within a particular class-interval as its **frequency.** In our example, we group together those whose ages are 25 years and more but less than 30 years

and place their number against the class interval 25-30 years group those who are 30 years and more but less than 35 years and place their number against the class interval 30-35 years and so on. Evidently the magnitude of our class interval is 5 years. Unit is the most common magnitude. So far as possible the magnitude of all class intervals should be uniform so that the labour of calculating different statistical constants may be minimised. Even if a particular class interval contains no frequency the class interval must be entered in its proper place otherwise errors might be made in plotting the results. The limits of the class intervals should preferably be so fixed that the mid point of each falls on an even unit and not on a fraction. One might ask—How many groups should there be? In answer it may be said that the number of class intervals is dependent on the nature of the inquiry. In general a number of groups in the neighbourhood of *20 is the most* satisfactory provided the number of observations is reasonably large. Thus classification according to class intervals is obtained when a numerical characteristic is considered and each group is sub-divided into a number of classes or groups rather arbitrarily. Table 1 on page 102 stands as an illustration.

The above class intervals *viz* 25-30 years, 30-35 years etc. are expressed according to **exclusive method,** that is the upper limit of the one class interval is the lower limit of the succeeding class interval. An item 29.99 years would fall in the class interval 25-30 years while an item exactly 30 years would be taken to the second class interval (30-35 years). This difficulty can be won over by classifying the class interval as 25 and under 30 years, 30 and under 35 years *and so on*.

Class intervals are also expressed by the **inclusive method**. The above class-intervals arranged according to the exclusive method would be expressed as 25-29 years, 30-34 years etc. according to the inclusive method. In this case the upper

limit of the one class-interval is also included in the class-interval itself. The first class would include all items between 24.5 and 29.5 years. To be still more unambiguous, the class-intervals may be expressed as 25-29.9 years, 30-34.9 years. But the inclusive method is not in general use since in adopting it the idea of continuity in the limits of class-intervals is lost.

Statistical Series.

If the quantities or values of some aggregate are measured, counted or weighed, or numbers in some group or class are counted, and they are placed one after another, the result is the statistical series. Briefly, a **statistical series** may be defined as things or their attributes arranged according to some logical and systematic order.

Time, Spatial and Condition Series.

There are three bases of classification of data; time, space and condition. Accordingly there are three types of statistical series: (1) **historical** or time (2) **spatial** and (3) **condition.** In the first, facts are arranged with respect to time, for instance, index numbers of wholesale prices of wheat in India over a period of time detailed in chronological order. In the second, the controlling factor in presentation is place: variations are noted geographically. Production of wheat in India for a given data arranged according to different provinces would constitute a spatial series. In the third, variations in size and amount of things or their attributes are shown. The different measurements of natural phenomena are usually distributed about a norm. If heights, weights, or ages of students, or lengths of a number of leaves chosen at random are measured, the different measurements, when arranged in logical order, shall constitute a condition series, and it will be noticed that though the different measurements would vary, a most common, predominant weight height, age or length of leaves would be found. We shall see

later (in Chapter X) that this 'most common' length is called the mode of the series Condition series take the form of what are called 'frequency tables', *e.g.*, table 1, page 102

Continuous and Discrete Series.

Series may be continuous or discrete When the items of a series are not capable of being determined with mathematical accuracy, but are always measured by approximation and can only be placed within certain limits the resultant record is a **continuous series.** Table 1 serves as an example On the contrary, where the items are exactly measurable and their record shows definite breaks between one value and the other succeeding it the resultant record is a **discrete** or broken or **discontinuous series.** For illustration see table 5, page 123

Measurements of weights magnitude and volume constitute continuous series for apparently there is no limit to the sub-division of maunds miles and gallons The numbers of labourers in factories, of spots in dice-throws or of pages in books form discrete series, since they must give integral numbers and are incapable of sub division.

TABULATION

After the data have been classified they may be tabulated, that is, put into tabular form **Tabulation stands for the systematic and scientific presentation of quantitative data in such a form as to elucidate the problem under consideration.** Its function is to arrange in an orderly manner the answers to those questions with which the inquiry is concerned Tables are intended to summarize the information obtained in course of an investigation

Rules and Precautions for Tabulation

Some precautions in drawing up tables are necessary The given data may be grouped in one table or several tables A single table shall, no doubt, bring the entire data into proxi-

mity; but if it is too large, it shall confuse the eye and lead to great difficulty in following the columns and rows at a glance. To do away with such inconvenience it may be broken up into several separate tables. Further, several comparisons of different nature should not be jumbled up in one table. Each table should be a unit. Usually there should be separate tables for different distinct purposes. Again there should be few main. divisions with several sub-headings under each. If the number of headings is very large, the main facts to be compared may not be adequately emphasized. Of course, the exact number of divisions and sub-headings shall be determined by the data in hand. Each table should be so complete in itself that it may not be made more intelligible by re-drafting. The table should suit the size of the paper on which it is drawn. So, the width of each column and row should be properly calculated and headings correctly arranged before the permanent table is ruled or figures are entered in it. Totals, averages, percentages and the numbers that are to be compared should be placed close together, and, if possible, they should be placed in the same vertical column rather than the same horizontal row. Columns that are to be compared should be placed adjacent to one another. The rulings in tables should be such that principal groups are separated by thick or multiple-ruled lines. Unimportant data may be grouped together and placed in 'miscellaneous' group. Items which are in any way different from the rest of the items—*e.g.*, estimated figures, revised figures—should be marked with an asterisk or number, and an explanatory note given beneath the table. The table should be given a suitable title. This title and the titles of sub-headings should be self-explanatory, so that no reference in the text or footnotes for the purpose may have to be looked up. The title should neither be too small nor ambiguous. The column-heading should indicate the unit used, as 'height in inches,' 'price in rupees', 'weight in tons'. Large digits may be approximated and mentioned in thousands,

lakhs or millions A table may show absolute figures, increase or decreases from past years' figures, percentages etc, according to the nature of comparison to be made All items should be carefully checked *before entering in the table or* totalling them up Arrangement should also be made in the table for testing a cross-checking There should be no over writing otherwise the neatness of the table would be lost A written analysis pointing out principal conclusions and possible errors with probable reasons for them, should accompany the table

Different types of Tabulation.

In very general terms tabulation may be distinguished a simple and complex A **simple table** contains data respecting one characteristic only, information relating to other characteristics being left out. Table 1 is a case of simple tabulation. A more **complex table** may contain figures relating to several characteristics, Tables 2, 3 and 4, represent this type

Tabulation is also classified as Single, Double, Treble and Manifold A **single tabulation** is one that answers one or more groups of independent questions The following table given the frequency distribution of marks obtained out of a maximum of 50 by the students of a class in their test in economics.

Table 1 *Frequency Distribution of Marks in Economics*

Marks group	Number of students (frequency)
0-5	4
5 10	6
10-15	10
15 20	16
20-25	12
25 30	8
30-35	4

The table is capable of furnishing a first appproximation to the answer to an inquiry into the 'ordinary marks' obtained by the students in the test. It tells, for instance that the number of students getting marks between 15 and 20 is the highest as compared with the number in any other group. A still simpler table will be one showing yearly variation of some characteristic, say, progress of cotton mill industry in India or of trading profits of a company.

Double tabulation shows the sub-division of a total according to two categories, and is capable of answering two mutually dependent questions. Table 2 is an illustration of this type of tabulation. It shows the distribution of 376 industrial disputes of the year 1941 into (i) different kinds of mills and (ii) different quarters of the year.

Table 2, *Industrial Dispute in India in 1941.*

For quarter ending	Industrial Disputes			Total
	Cotton & Woollen Mills	Jute Mills	Others	
31st March ..	30	1	40	71
30th June ..	52	3	66	121
30th September ..	34	3	41	78
31st December ..	33	10	63	106
Total ..	149	17	210	376

Treble tabulation sub-divides a total into three distinct categories and answers three mutually dependent questions. Table 3 is a blank table to illustrate treble tabulation. It shows the distribution of India's population into urban and rural, for main religions, in Provinces, and States and Agencies.

Table 3 *Distribution of India's population in urban and rural according to main religions in Provinces and States and Agencies*

Religion	Provinces			States & Agencies			Total		
	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Hindu									
Sikh									
Jain									
Buddhist									
Zoroastrian									
Muslim									
Christian									
Jew									
Tribal									
Others									
Total									

Manifold tabulation is one that divides a total into several categories generally more than three. The following blank table drawn to show the distribution of population in Indian Provinces according to age, sex, literacy and caste illustrates manifold tabulation.

Table 4. *Distribution by age, sex, literacy and caste in Provinces of India.*

Province	Caste	Age-Group	Male	Female	Total
Bengal	Kayasth	0-25			
		25-50			
		50-75			
		Over 75			
		Total			
	Brahmin	0-25			
		25-50			
		50-75			
		Over 75			
		Total			
	All Castes	0-25			
		25-50			
		50-75			
		Over 75			
		Total			
Bihar	Kayasth	0-25			
		25-50			
		50-75			
		Over 75			
		Total			

EXERCISES

(1) Define Classification and Tabulation and show their importance in statistical studies.

(2) In collection and tabulation common sense is the chief requisite and experience is the chief teacher —Bowley

What precautions in your opinion are necessary to avoid statistical errors in the collection and computation of economic data?

(M A Agra 1940)

(3) What different types of tabulation do you know? Indicate their characteristics

Draw up blank tables to illustrate your answer

(4) Explain the Temporal Spatial Qualitative and Quantitative bases of classification

(5) Point out the mistakes made in the following blank table drawn to show the distribution of population according to sex and literacy in five towns in the U P —

	Males					Females				
	Allahabad	Lucknow	Benares	Agra	Aligarh	Allahabad	Lucknow	Benares	Agra	Aligarh
Number of Literates										
Number of Illiterates										

(6) Re-arrange the following blank table with a view to make it more intelligible

Sex	Brahmin		Rajput		Kayastha		Harijan	
	Literate	Illiterate	Literate	Illiterate	Literate	Illiterate	Literate	Illiterate
Male								
Female								

(B Com, Alld 1940).

(7) Draw up in detail, with proper attention to spacing double lines, etc., and showing all sub-totals, a blank table in which could be entered the numbers occupied in six industries at two dates distinguishing males from females, and among the latter single, married and widowed.

(M.A., Alld., 1940).

(8) Prepare a specimen form in blank, with suitable heading and spacing, for use in collection of data on *one* of the following:—

(*a*) Survey of trades in your districts.
(*b*) Standard of living of middle class families in a small town.
(*c*) Expenses of students in a University.

(Dip. in Econ., Madras, 1931).

(9) Explain how you would tabulate statistics of death from principal diseases by sexes in different provinces of India for a period of five years.

(B.Com., Cal., 1937).

(10) Prepare a table with a proper title, divisions and sub-divisions to represent the following heads of information:—

(*a*) Imports of cotton piece goods in India.
(*b*) From U. K., Netherlands, Belgium, Switzerland, Italy, Straits Settlement, Japan.
(*c*) Amount of piece goods from each country.
(*d*) The value of goods from each country.
(*e*) Pre-war average, war average, Post-war average 1924-25, 1925-26, 1926-27, 1927-28.
(*f*) Total amount imported during each period.
(*g*) Total value of imports during each period.

(B.Com., Luck., 1930.

(11) Discuss the function and importance of tabulation in a scheme of investigation.

Prepare blank tables, showing the distribution of the student of a University according to age, class, and residence, for arranging (a) physical training, and (b) seminar classes.

(12) What precautions would you take in tabulating your data? Prepare a blank table to show the distribution of population according to sex and four religions, in five age-groups, in seven important cities of U. P.

(B Com, Agra, 1937)

(13) Draw up two independent blank tables, giving rows, columns, and totals in each case summarizing the details about the members of a number of families, distinguishing males from females, earners from dependents, and adults from children

(M A, Cal, 1935).

(14) What is a statistical series?

Differentiate between continuous and discrete series Give illustrations

Also distinguish between ordinary and cumulative frequencies

(15) Write short notes on

Frequency table, frequency, class limits, class interval, magnitude of class interval, inclusive and exclusive methods of classification treble tabulation, manifold classification

(16) What are the essentials of a good statistical table? What rules and precautions should be observed in drawing up a table?

(17) Following are the heights in inches of 53 students of a class Tabulate them by grouping them in class intervals of five inches —

58, 56 57, 57, 52, 53 56, 51, 49 48, 47, 48, 49,
60 62, 46 51, 60, 50, 53 54, 55, 56, 57, 56, 54,
59 60 47, 48 64 65, 63, 46, 62, 61, 52, 52, 53,
52, 55, 54 52 52, 53, 55, 48, 50, 51, 52, 66, 61,
52

(18) Prepare a blank table to show the value and quantity of different kinds of cotton goods imported into India from different countries of the world during the past six years

(19) Classify the following according to attributes —

(a) occupations in India, (b) exports and imports of India, (c) wants of university students, (d) books of a college library, (e) religions of India

(20) Following are the weekly earnings of labourers in a factory :—

Earnings Rs.	a.	p.	No. of labourers
4	6	0	25
10	10	6	6
5	9	3	35
5	1	0	42
6	11	9	30
7	12	3	21
9	9	0	17
8	2	0	15
10	0	0	8
5	0	0	52

Tabulate the above data in classes of Rs. 4-5, Rs. 5-6 etc., and give a suitable heading to the table.

(21) Draft tables to show the distribution of population by:

(*i*) age sex and civil condition,

(*ii*) sex and infirmities,

(*iii*) sex and occupations,

(*iv*) age, sex and literacy.

(22) How will you proceed to classify the observations made, and what points will you take into consideration in tabulating them? Mention the kinds of tables generally used.

(B.Com., Agra, 1941).

Rate per unit is called **statistical co-efficient.** If the birth rate is 30 per 1,000 the co-efficient is 03 The characteristic of this co-efficient is that if it is used to multiply a total, (*e g*, population) an allied number (number of births) would be obtained.

Purpose of Computing Statistical Derivatives.

Simple derivatives are computed to compare statistical groups In the computation of rate per cent, or rate per *mille* observations are reduced to a common denominator and comparison is thereby facilitated If in University A 900 candidates were successful out of 1200 who appeared, and in University B 980 passed out of 1400, a comparison of the absolute number of successful candidates—900 and 980—without considering the number of those who appeared at the examination would lead one to declare the result of the B University as better than that of the A. But, when the two results are reduced to a common denominator, say, expressed in percentages, this impression will be reversed The percentage of successful candidates in A is,

$$\frac{100\times\text{Number of successful candidates}}{\text{Number appeared}}=\frac{100\times900}{1200}=75$$

Similarly, the percentage in B is,

$$\frac{100\times\text{Number of successful candidates}}{\text{Number appeared}}=\frac{100\times980}{1400}=70$$

Both the results have now been reduced to a common denominator, 100 It is evident that the percentage of success in A University is higher than that in B University Therefore, A's result is better than B s The usefulness of relative numbers for purposes of comparison is thus clear

But relative numbers can also be used for another purpose, *viz*, computing the size of an unknown mass from a known one. The known mass *i e*, the relative number, may be an actual figure or an estimated one Relative numbers are often estimated when there are no sufficient data for their computation.

Such estimates can be employed to know the size of the mass to which they relate. Statisticians have often used them to obtain the population figures of past times. If we know, historically, the number of artisans or beggars of a city or country, we may make an estimate of the percentage of the entire population that the artisans or beggars probably formed, on an average, at the time under consideration, and hence compute the total population for that time. The estimation of the artisans or the beggars is made possible by the law of statistical regularity: the ratio between definite statistical masses is often fairly constant. This ratio can be easily estimated to be within certain limits. This holds good, for example, for the relationship between population and births and deaths. Every estimate, however, must be regarded as simply an approximate value. It may or may not be accurate. Therefore, the size of an unknown mass computed from an estimated relative number must also be regarded as merely approximate.

But, where actual percentages are known the mass or population to which they relate can be ascertained to a considerable degree of precision. For, given that the number of successful candidates at a certain examination was 900, and this constituted 75% of the total, it is easy to see that the total number of candidates who appeared at the examination was 1200.

Derivative Series.

A set of relative numbers or simple derivatives of the same kind spread, say, over a period of time would constitute a derivative series. The special feature of a derivative series is that it eliminates the factor or factors obstructing effective comparison: all figures are related to a common denominator and comparison is facilitated. A number of figures representing burden of income-tax per head of population over

a period of time forms a derivative series. The series would eliminate the main effects of demographic changes. Population on which the burden is computed may change, yet the burden per head of population for one year shall be comparable with similar burden for another year. The actual figures of population and amount of tax would not be so easy to compare from year to year because of their fluctuations.

The test of a derivative series lies in its stability, which is determined by measures of dispersion. These measures shall be discussed later (in Chapter XI), but it will be useful to note here that higher the degree of stability, greater is the reliability of a derivative series. To attain stability some simple precautions should be kept in view in computing simple derivatives.

Rules and Precautions for computing Derivatives

Both the computation and the use of relative numbers need caution. The rule should be to procure as much homogeneity in the data as possible. For example, general death rate for a town or a country may be computed by multiplying the number of deaths by 1,000 and dividing the product by the total population, but this general death rate, or crude death rate as it is called, relates to heterogeneous mass, since deaths vary with age and sex compositions of the population. We shall later see (in Chapter X) how this defect can be remedied.

In order that there may be no misunderstanding, the basis of calculation of the relative numbers should always be given. If we are told that the price of a commodity increased 10 per cent, decreased 15 per cent, increased 25 per cent, decreased 20 per cent, and then increased 15 per cent over a period of time, it would be difficult to say what exactly the change over this period was. If the changes were based on the original price it would be found that the change over the whole period was 15% on that price. If, however, the changes were based

on the prices ruling at the time of each particular change, the change over the period would be found to be about 7.5% of the basic price. If the basis on which percentages were calculated were known this variation in result would not occur.

Again, percentages, or other relative numbers, should be used only when the factors which are to be expressed in percentage form are themselves comparable. If a company whose issued and paid up capital was Rs. 150,000 earned profits at a uniform rate of Rs. 15,000 per year for five years, its percentage of profits to capital would be 10 for the period. Suppose the company increased its capital to Rs. 250,000 in the sixth year and the profits increased from 15,000 to Rs. 22,500 in that year, the percentage of profits to capital would be 9 only. Then, if a table showing only the percentage profits were prepared for the six years, it might lead one to conclude that profits had declined in the sixth year while, in fact, the actual profits had increased. In such cases it is advisable to show the amount of capital over the different years, the total profits earned from year to year and in the last the rate of percentage profits. This would avoid all fallacy.

Lastly, percentages should not ordinarily be used when the number of items in one of the series to be compared is less than one hundred. Similarly, rates per thousand or rates per ten thousand should not be computed when the number of items is comparatively very small. Advertisements very often appear in the newspapers: "Join X school. This year's results 100 per cent." Another institution may have a percentage of only 92. A prospective candidate may be led to think better of the first institution. But, if on an inquiry it is found that only 3 candidates appeared from X school and all got through, while 250 candidates appeared from the second institution of which 230 came out successful, opinion will have

to be reversed for it is always more difficult to get the same percentage result from a much larger number Mathematically both the percentages are absolutely correct but *statistically* the percentage relating to the result of X school is not significant It is therefore again established that comparison through percentages alone is not sufficient unless the data on which they are based are homogeneous and capable of comparison

Ratios

To eliminate the chances of fallacious conclusions which might result from the use of percentages when their bases of calculation are not specified some statisticians emphatically recommend the use of ratios in place of percentages Then we shall say that the price of the commodity increased in the ratio of 100 110 rather than that it increased 10%

But ratios must also be used with caution otherwise wrong inferences might be drawn Suppose 800 candidates appeared at a certain examination of which 600 came out successful The ratio of passes to failures is therefore 6 2 Further supposing college A *coached* 500 out of the 800 candidates and of these 400 passed so that the ratio of the successful to the failed candidates is 4 1 whereas of the remaining 300 candidates 200 must have passed which gives the ratio of successes to failures as 2 1 It might appear that college A achieved twice as good results as all those colleges taken together through which the remaining 300 candidates appeared This conclusion is fallacious as it is not known whether all the 300 students were given adequate coaching or they simply *appeared* through certain colleges If they were not properly coached they did not stand the same chances of success as those 500 candidates who were given due coaching It is advisable, in such cases to show the ratios of all the coaching institutions in order to ascertain which one of them was really the best

Use of Simple Derivatives.

Ratio, rates per unit, per hundred, per *milles* are widely used and easily understood. Sex ratio, cost per unit of output, income *per capita*, percentage rate of interest, percentage of exports or imports to total rade, birth and marriage rates per thousand, mortality rate per ten thousand, burden of tax per head of population, net reproduction rate—these derivatives suggest the variety of field to which they are applied. They are very commonly used in business. social and administrative statistics.

EXERCISES

(1) Define a statistical derivative, and point out the usefulness of its computation in statistical studies.

(2) Clearly distinguish between Subordinate and Co-ordinate derivatives.

(3) What purposes do statistical derivatives serve? Do they give the whole information about the series from which they are derived.

(4) What is a derivative series? How does it differ from (*i*) series of individual observations and (*ii*) frequency distribution?

(5) What rules and precautions will you observe in computing percentages, and why?

(6) What are ratios? Why are they considered as better than percentages?

(7) What precautions are necessary in using ratios and percentages?

(8) Explain what you understand by (*a*) amount of tax per tax payer, (*b*) burden of tax per head of population, (*c*) net reproduction rate, (*d*) income *per capita*, (*e*) yield per acre, (*f*) cost per unit of output.

(9) Write a note on the importance of simple derivatives to businessmen, professional speakers. legislators and layman.

(10) Point out the ambiguity or mistake, if any. in the following statements:—

(1) The death rate in the American navy during the Spanish American war was nine per thousand while in the city of New York for the same period it was sixteen per thousand It was safer, then to be a sailor in the American navy than to live in New York City

(2) 13% of the total population in 1941 in India was urban as against 11% in 1931 Therefore, the number of towns in India considerably increased during the decade

(3) Population of India increased 15% in 1941 over the population in 1931 Therefore the consumption of food grains per head fell in 1941

(4) The increase in the wage of a labourer was 20% Then the wage decreased 25% and again increased 15% Therefore the resultant increase in the wage was 10%

(5) Cows are multiplying faster than human beings in India The consumption of milk by human beings is therefore increasing

(6) 50 candidates appeared from a college in the B A examination of which 60% were successful and two obtained first division From another college 5 candidates appeared at the same examination and 80% passed none being placed in the first division The latter college showed better result than the former

(11) The following table shows the growth of India's population as recorded by successive census —

Year	Population (000 000's omitted)
1872	210
1881	250
1891	290
1901	295
1911	315
1921	320
1931	353
1941	389

Calculate the percentage increase for each successive year over the preceding year's population

CHAPTER X

STATISTICAL AVERAGES

Simple statistical derivatives, by themselves, are insufficient to give a summary description of the peculiarities of a series, nor can they be used as types representing the series. They throw light only on the relative aspects of a series, and are not characteristic of the data. Therefore, some other method of precisely and concisely describing the series has to be devised. This is the method of statistical averages. Through it a number, representative or characteristic of the entire group, is computed, which affords the central idea of the series and can be used in place of the data.

Statistical Average and its Object.

Averaging is the process of condensation. **An average is a single simple expression in which the net result of a complex group or large numbers is concentrated.** It is a representative number of the group, its gist. An average brushes off the irregularities of a series, levels all differences of the individual items and presents complex data and unwieldy numbers in a few significant figures. It thus gives a bird's-eye view of an aggregate, and can be substituted for individual items in further calculations regarding the series. It is a typical item representing the group and is, therefore, also called **type.** As such, it describes the group or throws light on its characteristics better than any other value in the group.

There are two main objects of computing a statistical average. One is **to give a concise picture of a large group,** to describe the series it represents. The other, which follows from the first, is **to afford a basis of comparison with other**

groups It follows that an average may be computed for its own sake or as a means to another end which may be comparison or measurement of dispersion[1] or of skewness[2] of the series This is quite obvious It is difficult to grasp the significance of the ages of the inhabitants of a country if we are given the age of every person in the country but the average age of the people of the country is something definite and intelligible Similarly two series each containing ages of different people in two countries even if compressed into a few magnitude classes will not afford a comparison between the ages of the people in the two countries If however some sort of average of the two series is computed average ages in the two countries shall be comparable at a glance And comparison of the average ages shall usually be equivalent to comparison between the two series

Homogeneity of Data.

It is necessary to say that the data from which averages are computed must be as largely homogeneous as possible For if the different items are not alike in relevant aspects there is no sense in grouping them together, and consequently, no justification for computing their averages The comparison of only those averages yield reliable conclusions which refer to homogeneous masses The comparison of averages computed from heterogeneous series may easily mislead and is only reliable under certain special conditions The significance of averages lies in the fact that they exhibit the result of the activity of complex causes in one characteristic figure The average wage for instance of a given group of workers gives a measure of the factors determining wages in that group It is important that this average should refer to as unified complex of causes as possible since then alone will it be reliable

---

For Dispersion and Skewness see Chapter XI

for purposes of comparison or as a type. If the wages in two factories are determined by quite different causes, the average wages for all the workers in the two factories shall not yield a trustworthy comparison. Hence the importance of homogeneous series.

Homogeneity can be attained by (1) eliminating the unlike from the like items and (2) dividing the like items into groups as nearly homogeneous as possible. In our example, it will be necessary to disregard the cases in which workers for personal reasons do not receive any wages at all. Among those getting wages, workers may be distinguished whose wages are influenced by different and independent causes. They will have to be divided into more homogeneous parts, so that the averages may refer to an unified complex of causes, and thus be reliable for comparison between two periods or two places. Our wage-earners may consist of males and females, the latter getting lower wages. Therefore, male workers will have to be separated from female workers, and separate averages computed. There is still a possibility of these two groups being sub-divided into more homogeneous parts, *e.g.*, skilled male workers and unskilled ones. The extent to which homogeneity should be attained shall, however, be determined by the purpose for which averages are required. We hear of the average tax per tax-payer as also of average tax per head of population. Both have different purposes: the first is a measure of the burden on the tax-payers only, the second of that on the whole population.

Kinds of Average.

The following four kinds of average, or mean, are in common use.

(1) The Mode, (2) The Median, (3) The Arithmetic Average, and (4) The Geometric Mean.

In addition there are other forms of average such as the Harmonic Mean and the Quadratic Mean but they are not in common use

THE MODE

Mode is the value of that item in a variable which occurs most frequently, or is repeated the greatest number of times It lies at the position of greatest density It is the typical measurement or most fashionable point It is the usual and not casual size of item in a series When we speak of the average student the average wage the average rent etc we generally imply the modal student the modal wage the modal rent If we say that modal marks obtained by students in a class are 40 we mean that 40 are the predominant marks, i.e., the largest number of students secured 40 marks As *high as 70 marks and as low as 15 marks are* exceptions they are much less frequented they are non modal

Since mode is the most frequent size it appears it is easy to locate it Really it is so if there is a single well defined mode Then the size of item or the group containing the maximum frequency can be easily located in a frequency table. But it is not improbable that there may be numerous irregularities in the table so that the position of the mode would become indefinite and modal size would not be easy to locate In such cases frequencies are adjusted by the process of grouping i.e., by widening the groups into which the frequencies fall until a modal size of item or a modal group presents itself This modal size is the mode in a discrete series while in a continuous series the mode will be located by *interpolation* in the modal group on the assumption that the frequencies of the groups on either side of the mode influence the mode in proportion to their respective numbers

Location of Mode: Discrete Series.

Example 1. Required to find the mode in a discrete series.

Table 5. *Location of Mode by Grouping.*

Size of item m	Frequency f (1)	(2)	(3)	(4)	(5)	(6)
4	2	7		15		
5	5		13		22	
6	8	17				29
7	9		21	35		
8	12	26			40	
9	14		**28**			**43**
10	14	**29**		**40**		
11	**15**		26		39	
12	11	24				33
13	13		22	29		
14	9	16			20	
15	7		11			14
16	4	7				
17	3					

The frequencies given in column (1) are first grouped by two's in columns (2) and (3), and then by three's in columns (4), (5) and (6) and the maximum frequency in each column is indicated in heavy type. But no fixed point where frequency may be the largest is obtained; the mode seems to change with change in the grouping. According to column (2) it may be 10 or 11, while according to column (5) it may be 8, 9 or 10. The following table shows the sizes of maximum frequency in different columns.

Table 6 Analysis Table

Column	Size of item containing maximum				
(1)				11	
(2)			10	11	
(3)		9	10		
(4)			10	11	12
(5)	8	9	10		
(6)		9	10	11	
No of times	1	3	5	4	1

From the above table we find that 10 is the size of item which is most frequented It is not true of any other size The mode is therefore, located at 10

A glance at the frequencies in column (1) table 5 might lead one to think that size 11 is the mode since it contains the largest frequency in that column But this impression is corrected by the process of grouping which clearly shows that mode is influenced by the frequencies of the neighbouring sizes *It is therefore evident that it is not always easy to locate the mode by mere inspection* Inspection can give reliable results only where the frequencies run fairly regularly and mode is unquestionably clear and well defined We shall see it in the following example

Location of Mode Continuous Series.

Example 2 Required to locate the mode in a continuous series whose frequency distribution is given in table 1 page 102

We find that the frequency distribution in the said table is fairly regular and the 15-20 marks group indisputably contains the maximum frequency. Without any preliminaries we can say that the mode lies in 15-20 marks group. If we group the data, first in 10 marks group and next in 15 marks group, and then prepare from the frequency columns an Analysis Table as we did in example 1 above, we shall arrive at a similar conclusion. Having known the class containing the maximum frequency we shall locate the mode in that class on the assumption already noted, *viz*, according to the weights or influence of the neighbouring groups.

The formula for it will be as follows[3]:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

where, Z stands for the mode,[4]

l_1 and l_2 stand for the lower and upper limits of the modal group,

f_1 stands for frequencies in the modal group,

f_0 „ „ „ next lower group,

f_2 „ „ „ next higher group.

Applying the above formula we have,

$$Z = 15 + \frac{16 - 10}{32 - 10 - 12}(20 - 15)$$

$$= 18 \text{ marks.}$$

---

[3] This formula, it is held, is more accurate than the customary formula, *viz.*

$$Z = l + \frac{f_2}{f_2 + f_0}(l_2 - l_1)$$

[4] In all our further calculations mode will be represented by Z.

The process of grouping is in fact a method of smoothing out the irregularities of the series and can be profitably employed even in a series which does not appear to be irregular. Rather, it should be resorted to in all elementary work.

In exceptional cases where the distribution of frequencies is very irregular two or more groups on either side of the modal class interval may be used as weights. But recourse should not be had to it if the series is multi modal. It should be noted that all series do not possess a single or even a well defined mode. Some are bi modal, some tri modal i.e. have more than one mode while others can hardly be said to possess a mode at all. Therefore efforts should not be wasted over forcing the appearance of an exact mode when in fact one does not exist. True mode should not be expected in a series which is *markedly* asymmetrical.[a]

Advantages of the Mode

1. It is easily understood and has a general and precise usage.

2. It eliminates extreme (and therefore abnormal) variations. That is its value is not affected by stray items differing much from it in their values.

3. For its *determination it is not necessary to know* the extreme items except that they are few. Only the size of the middle items need be known.

4. It refers to a measurement whose expectation in the series is the greatest. That is it is the most likely and not isolated example.

5. It can be located by mere inspection in certain cases.

Disadvantages of the Mode

1. It is frequently ill-defined and indefinite. A modal size of city may convey any meaning.

[a] For precise meaning of asymmetrical series see Chapter XI

2. It is often indeterminate and, therefore, difficult to locate.

3. It is incapable of being located by any simple arithmetical process.

4. It rejects all exceptional instances and is, therefore, not useful in those cases where weights are to be given to extreme variations.

5. It may not be fully representative of a group in which items of uniform size are comparatively small. For instance, if in a community of 200 people only 4 people earned Rs. 30 each while the earnings of the rest were at any figure other than Rs. 30, and no other four people received an equal amount, Rs. 30 would be the modal earnings simply because they were earned by the maximum number of people. This difficulty, however, can be got over by using class-intervals of considerable magnitude.

6. It is unsuitable for further algebraical treatment.

7. Mode multiplied by the number of items does not yield the total value of the items.

Uses of the Mode.

The concept of mode is readily intelligible, and is applied in many cases in daily routine, though involuntarily. We often hear people say 'Average calls on my telephone are 15 a day', 'Average size of shoe sold at my shop is such and such', 'The Average page contains 300 words', 'The average student spends Rs. 50 a month'. In all such cases what they mean by the average is really mode, that is, the likeliest figure. If we are required to guess the average of a certain phenomenon, we shall generally, and rightly too, guess the mode, the dominant or prevailing size.

The use of mode is now increasing in business. It serves as a reliable guide in business forecasting. It is being realized

that it is of great value in studying output Modal output per machine can be ascertained by sorting the output of similar machines and finding the output which is more or less the same over a period of time If this modal output is left far behind in subsequent year reasons for it may be traced back to defects in machines or their handling, lack of skill on the part of operatives or any other inefficiency This useful work might remain undone or its urgency may not be felt, in the absence of a knowledge of the modal output which acts as the standard for comparison Similarly modal time for producing a commodity may be ascertained which would stand for the most likely time that would be required to turn out similar goods under similar conditions. On the basis of this modal time cost of producing a certain number of commodities may be calculated Mode thus has great potentialities for being employed in business and commerce profitably

Meteorological forecasts, which are proving very important to mercantile and other interests, are really based on the mode

THE MEDIAN

Median is the value of that item in a series which divides the series into two equal parts, one part consisting of all values less, and the other all values greater than it. That is, if a series is *arrayed* or, which comes to the same thing, the values of its items are placed side by side in ascending or descending order of their magnitude, the value of the middle item of the array is the median. If the students of a class, 43 in number, be asked to stand in order of their height, the 22nd student from either side shall be the one whose height will be called the median height of the class This method of picking up the median item can be symbolically expressed as follows

$$M = \text{Size of} \left(\frac{n+1}{2}\right)^{\text{th}} \text{item}$$

where, M represents the median[6], and n the number of items.

Determination of Median: Individual Measurements.

Example 3. Required to find the median in a series of quantitative individual observations, relating to the monthly expenditure incurred by 35 students in a boarding house.

The given figures are first arrayed as follows:

Table 7. *Monthly Expenditure of 35 Students arranged in Ascending order of Magnitude.*

Serial No.	Expenditure Rs.	Serial No.	Expenditure Rs.	Serial No.	Expenditure Rs.	Serial No.	Expenditure Rs.	Serial No.	Expenditure Rs.
1	35	8	40	15	45	22	46	29	50
2	35	9	41	16	45	23	47	30	50
3	36	10	42	17	45	24	47	31	52
4	38	11	42	18	45	25	47	32	52
5	38	12	44	19	46	26	48	33	54
6	40	13	44	20	46	27	48	34	55
7	40	14	44	21	46	28	48	35	60

[6] Median is the *size* or *value* of the middle item, and *not the rank* or *the number* of such item.

In our further calculations M will represent the median.

F. 9

Applying the above formula we have,

$M=$ Size of $\left(\frac{n+1}{2}\right)^{th}$ item; n, in this case, equals 35.

$=$ Size of $\left(\frac{35+1}{2}\right)^{th}$ item,

$=$ Size of 18th item,

$=$ Rs. 45.

The number of items in the above example was odd and, therefore, there was no difficulty in locating the middle item (18th in this case). But, the number may be even. In such a case, the median is intermediate between the values of the two middle items. Suppose, in the above example, an additional, 36th, student's expenditure was Rs. 61. Then,

$M=$ Size of $\left(\frac{n+1}{2}\right)^{th}$ item; n, in this case equals 36,

$=$ Size of $\left(\frac{36+1}{2}\right)^{th}$ item,

$=$ Size of 18.5^{th} item,

$=\frac{\text{Size of } 18^{th} \text{ item} + \text{Size of } 19^{th} \text{ item}}{2}$

$=$ Rs. 45-8-0

Determination of Median: Discrete Series.

In a discrete series also the size of $\left(\frac{n+1}{2}\right)^{th}$ item shall be the median.

Example 4. Required to locate the median of the data given in table 5, page 123.

Table 8. *Cumulative Frequency Table.*

Size of item m	Frequency f	Cumulative Frequency cf
4	2	2
5	5	7
6	8	15
7	9	24
8	12	36
9	14	50
10	14	64
11	15	79
12	11	90
13	13	103
14	9	112
15	7	119
16	4	123
17	3	126

M = Size of $\left(\frac{n+1}{2}\right)^{\text{th}}$ item; n equals 126.

= Size of $\left(\frac{126+1}{2}\right)^{\text{th}}$ item,

= Size of 63.5^{th} item,

= 10.

[It should be noted that in this series the size of all items beyond the 50th and up to the 64th is 10.]

Determination of Median: Continuous Series.

In the case of continuous frequency distribution, however, the median will have to be interpolated in the class-interval containing the median, if the original data from which the given frequency distribution is derived are not available. *Interpolation* gives only an approximate value. It is done on the assumption that the magnitude of the median class-interval is uniformly spread over its frequency.

Example 5 Required to locate the median in the continuous frequency distribution given in table 11 page 102

Table 9 *Cumulative Frequency of Marks of 60 Students in Economics*

Marks Group m	Frequency f	Cumulative Frequency cf
0—5	4	4
5—10	6	10
10—15	10	20
15—20	16	36
20—25	12	48
25—30	8	56
30—35	4	60

$M = \text{Size of } \left(\frac{n+1}{2}\right)^{th} \text{ item} \quad n \text{ equals } 60$

$= \text{Size of } \left(\frac{60+1}{2}\right)^{th} \text{ item}$

$= \text{Size of } 30.5^{th} \text{ item}$

If we had the original data with us the size of 30.5th item could have been *directly* determined But since we don't have the data we can only *estimate* the median. 30.5th item is situated in the 15—20 marks group This group's frequency is 16 and magnitude 5. It is assumed that these 5 marks are evenly distributed over the 16 students The 20th student gets approximately 15 marks Therefore the 30.5th student shall get $\left\{\frac{5}{16}(30.5-20)\right\}$ or about 3.28 marks more than the 20th student Thus the size of 30.5th item in our series that is the median is 18.28 marks

The above calculation can also be symbolically expressed as:

$$M = l_1 + \left\{ \frac{l_2 - l_1}{f_1} (m - c) \right\}^{7}$$

where, M represents the median; l_1, the lower limit of the group in which median is situated; l_2, the upper limit of median class; f_{12} the frequency of median class; m, the number of middle item or $\left(\frac{n+1}{2}\right)$th item; and c, the cumulative frequency of the group lower than the one in which median is situated.

Applying the above formula we have,

$$M = 15 + \left\{ \frac{20-15}{16} \times (30.5-20) \right\}$$

$$= 15 + \left(\frac{5}{16} \times 10{\cdot}5 \right)$$

$= 18.28$ marks, approximately.

Advantages of the Median.

1. It is easily understood.

2. It eliminates the effect of extreme (and therefore abnormal) variations.

3. It can be determined without a knowledge of the magnitude of extreme items, provided the number of items is known.

4. It is usually, *e.g.*, when found exactly, an actual example from the data.

5. It can be located by inspection in certain cases.

6. It can be exactly located (*cf* Mode, which cannot be exactly located in a multi-modal or non-modal series.)

---

[1] Formulæ slightly different from this are also given by certain authors. We, however, feel that this formula is satisfactory, as it is in keeping with the assumption we have made for interpolation.

7 It is specially useful for considering data the items of which are incapable of being quantitatively measured A group of students may be made to stand in order of their intelligence The middle student shall represent the median intelligence Median can, therefore, be employed to serve as an average, yielding a sufficiently reliable representative, in an estimate of qualities like honesty, health, virtue, which cannot possibly be expressed in specific units

Disadvantages of the Median.

1 The fact whether the median is representative of the variable depends upon the nature of the distribution of the values It may not be representative when the distribution is irregular *i e*, when the items vary greatly in magnitude Let the runs made by the players of a cricket team be 0 1, 3 6 9 12 48, 60 60 60, 98 The median runs *12* are made only by one player while 60 are scored by three players Median is not a typical representative in this case

2 It cannot be *precisely* determined when it falls between two values Then it can only be *estimated* When estimated it may be a value not found in the series If in a class of 40 students, 20 secure marks varying between 10 and 20 and the other 20 secure marks varying from 25 to 35 the median marks would be indeterminate They would be assumed to fall between 20 and 25 which are not obtained by any of the 40 students. The median marks would give a fictitious number

3 It is not capable of being located by any simple mathematical process

4 It is not useful in those cases where large weight is to be given to extreme items for it treats all frequencies alike

5 It is unsuitable for arithmetic or algebraic manipulation

6. The aggregate value of items cannot be obtained when the median and the number of items are known.

7. It requires the data to be arrayed before it can be determined—an operation which involves considerable work.

Uses of the Median.

Median is easy to understand and is, therefore, useful for practical purposes. It is not only useful for the study of problems whose objects are not quantitatively measurable, but is also valuable in comparing such data as are difficult to measure individually and have to be grouped within certain limits. It is, therefore, of immense use in considering social phenomena like wages, distribution of wealth, skill, etc. The median, however, is not very suitable for being used in commerce, because commercial data are very widely dispersed *i.e.*, they are not highly regular in distribution. Where such is the case, we have seen, median is not a good representative. What the businessman has in mind is usually the mode.

QUARTILES, DECILES & PERCENTILES

The principle according to which median is determined can be extended to divide a series into any number of parts. The values of the items dividing a series into four equal parts are called **Quartiles.** When a series is arrayed and the median divides it into two halves, each of the lower and the upper halves can also be divided into two equal parts. The value of the item dividing the lower half is called the **First Quartile** or the **Lower Quartile** represented by Q_1, and the value of the item dividing the upper half is called the **Third Quartile** or the **Upper Quartile** represented by Q_3, median being the Second Quartile. A series is, thus, divided into four equal parts at the first and the third quartiles, and the median.

Similarly a series may be divided into ten equal parts In doing so we shall get nine dividing positions the values of which are called **Deciles.** We have, thus, nine deciles in a series the fifth decile being the median

Again a series may be divided into 100 equal parts giving ninety nine dividing positions, the values of which are called Percentiles There are, thus, 99 percentiles in a series the fiftieth percentile is the fifth decile, or the median

In similar manner we can have Quintiles and Octiles according as our series is divided into five equal parts or eight equal parts

Location of Quartiles, Deciles and Percentiles.

The principle of locating the median is the principle followed here also The given series is first arrayed Then if the series is composed of quantitative individual observations or is a discrete one the following formulæ shall apply —

Q_1 = Size of $\left(\frac{n+1}{4}\right)^{th}$ item,

Q_3 = Size of $\left(\frac{3(n+1)}{4}\right)^{th}$ item,

D_1 = Size of $\left(\frac{n+1}{10}\right)^{th}$ item,

D_2 = Size of $\left(\frac{2(n+1)}{10}\right)^{th}$ item similarly, for the rest of the deciles,

P_1 = Size of $\left(\frac{n+1}{100}\right)^{th}$ item

P_2 = Size of $\left(\frac{2(n+1)}{100}\right)^{th}$ item, similarly for other percentiles,

where Q_1 stands for first quartile[a], Q_3 for third quartile[a], D_1 for first decile D_2 for second decile P_1 for

---

[a], [a] In our further calculations Q and Q' shall stand for the first and the third quartiles respectively
Quartiles as also deciles and percentiles refer to the size of item and not to the rank.

first percentile, P_2 for second percentile, and n for number of observations.

Individual Observations.

Example . In the series given in table 7, page 129,

Q_1 = Size of $\left(\frac{n+1}{4}\right)^{th}$ item : n equals 35,

= Size of $\left(\frac{35+1}{4}\right)^{th}$ item ;

= Size of 9^{th} item,

= Rs. 41.

Q_3 = Size of $\left(\frac{3(n+1)}{4}\right)^{th}$ item : n equals 35,

= Size of 27^{th} item,

= Rs. 48.

D_4 = Size of $\left(\frac{4(n+1)}{10}\right)^{th}$ item : n equals 35,

= Size of 14.4 item.

= Size of 14^{th} item + $\frac{4}{10}$ (Size of 15^{th} item—Size of 14^{th} item)

= Rs. $\left[44 + \frac{4}{10}(45-44)\right]$

= Rs. 44.4.

P_{90} = Size of $\left(\frac{90(n+1)}{100}\right)^{th}$ item ; n equals 35,

= Size of 32.4^{th} item,

= Size of 32^{nd} item + $\frac{4}{10}$ (Size of 33^{rd} item—Size of 32^{nd} item)

= Rs. $\left[52 + \frac{4}{10}(54-52)\right]$

= Rs. 52.8.

Discrete Series.

Example 7 In the data given in table 8 page 131

$$Q_1 = \text{Size of } \left(\frac{126+1}{4}\right)^{\text{th}} \text{ item}$$

$$= \text{Size of } 31.75^{\text{th}} \text{ item}$$

$$= 8$$

$$Q_3 = \text{Size of } \left(\frac{3(126+1)}{4}\right)^{\text{th}} \text{ item}$$

$$= \text{Size of } 95.25^{\text{th}} \text{ item}$$

$$= 13$$

The various deciles and percentiles can also be determined in like manner

Frequency Distributions

If the data are grouped into certain defined limits quartiles deciles and percentiles shall be located by *interpolation* which yields approximate values The formulæ to be used shall be

$$Q = l + \left\{ \frac{l_2 - l_1}{f} (q - c) \right\}$$

$$Q_3 = l + \left\{ \frac{l_2 - l}{f} (q_3 - c) \right\}$$

where q and q_3 stand for first and third quartile numbers respectively and other symbols for what they did in interpolating the median except that the class intervals referred to shall be those relating to Q and Q_3 and not to median Similarly formulæ for interpolating deciles and percentiles can be framed

Example 8 In the continuous frequency distribution given in table 9 page 132

$$q = \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item} = 1[illegible]^{\text{th}} \text{ item}$$

$$q_3 = \left(\frac{3(n+1)}{4}\right) \text{ item} = 45.75^{\text{th}} \text{ item}$$

$$\text{Therefore, } Q_1 = 10 + \left\{ \frac{5}{10} \times (15.25 - 10) \right\} = 12.625 \text{ marks,}$$

$$\text{and} \quad Q_3 = 20 + \left\{ \frac{5}{12} \times (45.75 - 36) \right\} = 24.0625 \text{ marks.}$$

Characteristics and Uses of Quartiles, etc.

The quartiles, deciles and percentiles are not averages in the sense median is, since they refer not to the whole variable but only to parts of it. Of course, for determining them the part to which they relate is treated as the whole series. Thus quartiles are, in a sense, equivalent to the medians of the lower and the upper halves of a series, but they cannot be considered as averages of the first order, *i.e.*, as *sizes which can be taken* as types or substitutes of the whole series.

Yet, quartiles etc. give valuable information regarding the series. They indicate the distance within which certain parts of the series lie. Thus, knowing the quartiles of the data given in table 7, page 129, we may say that the middle half of the series lies between Rs. 41 and Rs. 48. Deciles and percentiles can also similarly yield the information characteristic of them. We shall refer to the importance of quartiles again, while considering the manner in which items in a series are distributed.[10]

THE ARITHMETIC AVERAGE

The arithmetic average, also called the arithmetic mean, is the quantity obtained by dividing the sum of the values of the items in a variable by their number. Thus, it is the average of common speech, an average quite familiar to the layman.

[10] See Chapter XI.

The term arithmetic average as defined above is applied to Simple Arithmetic Average in which all items are treated alike each item being considered only once, as distinguished from Weighted Arithmetic Average in which all items are not treated alike each item being assigned a weight in proportion to its importance in the series

There are two methods of computing the simple arithmetic average the direct and the short cut.

Computing Simple Average Individual Observations

In following the **direct method** the values of all the items in a given series are totalled up the total being called the Aggregate or Summation represented by the Greek letter Σ (Capital Sigma) The quotient obtained by dividing the summation of measurements by the number of items is the required simple arithmetic average

Symbolically

$$a = \frac{\Sigma m}{n}$$

where a represents arithmetic average[1] Σm represents summation of measurements and n represents the number of items

Example 9 Required to find the simple arithmetic average of the data given in table 7 page 129 by the direct method

$$a = \text{Rs } \frac{1580}{35}$$
$$= \text{Rs } 45\ 14$$

It is not necessary that the values of the items should first be arrayed as they are done in table 7 The original data

[1] In all our further calculations a will stand for the arithmetic

recorded as they occurred, could have been equally well utilized to find the aggregate.

The above method of computing the simple arithmetic mean involves considerable work of addition when the series is large and digits in each number are several. To save time and labour, the **short-cut method** can be utilized if the values of the different items happen to be nearly the same. To use this method, any size of item may be assumed as the average. Deviations of the value of each item from this assumed average should then be found and put down with proper algebraic sign. The algebraic sum of these deviations divided by the number of items when added to the assumed arithmetic average gives the required simple arithmetic average.

Symbolically,

$$a - x = \frac{\Sigma d_x}{n}$$

$$\text{or, } a = x + \frac{\Sigma d_x}{n}$$

where, a stands for the arithmetic average; x, for the assumed average; Σd_x for the summation of deviations from the assumed average; and n, for the number of items.

Deviation from the assumed average is calculated by subtracting the assumed average from the size of item, that is by the formula,

$$d_x = (m - x)$$

where, d_x is the deviation from assumed average; m, the size of item; and x, the assumed average.

The deviation, d would be positive or negative according as the size of item, m, is greater or less than the assumed average, x.

Example 10. Required to find the simple arithmetic mean of the data given below by the short-cut method.

In computing a by the direct method we shall be concerned with columns (1) (2) and (3) while in calculating it by the short-cut method we shall be concerned with columns (1), (2), (4) and (5)

Direct method $a = \frac{\Sigma m}{n} = \frac{1318}{126} = 10.46$

Short-cut method $a - r = \frac{\Sigma d_x}{n}$

$$a = 10 + \frac{58}{126} = 10.46$$

Computing Simple Average Continuous Series

When frequency distribution of a continuous type is given arithmetic average can only be calculated on the assumption that the values of all the items in each class are identical with the mid value of the class interval Both the direct and the short cut methods can be followed After finding out the mid values of the class intervals the procedure of calculating the arithmetic average would be the same as that in the case of discrete series.

Example 12 Required to find the simple arithmetic average of the data given in table 1 page 102 by the direct and the short-cut methods

In table 12 we shall not be concerned with columns (5) and (6) for the direct method and with column (4) for the short-cut method

Table 12. *Calculation of Simple Arithmetic Average of Marks of 60 Students by the Direct and the Short-cut Methods.*

1	2	3	4	5	6
Marks group.	Mid-value m	Fre-quency f	Total value of items (col. 2 × col. 3) mf	Deviations from assumed mean (17.5) d_x	Total deviations (col. 3 × col. 5) fd_x
0—5	2.5	4	10	—15	—60
5—10	7.5	6	45	—10	—60
10—15	12.5	10	125	— 5	—50
15—20	17.5	16	280	0	0
20—25	22.5	12	270	+ 5	+60
25—30	27.5	8	220	+10	+80
30—35	32.5	4	130	+15	+60
		$n=60$	$\Sigma m=1080$		$\Sigma d_x=+30$

Direct method: $a=\frac{\Sigma m}{n}=\frac{1080}{60}=18$ marks.

Short-cut method: $a-x=\frac{\Sigma d_x}{n}$

$\therefore\quad a=17.5+\frac{30}{60}=18$ marks.

In the foregoing examples on simple arithmetic average both the direct and the short-cut methods have been demonstrated. It will be seen that the answers in each case by both the methods are exactly the same. The saving in labour resulting from the use of the short-cut method is quite obvious. In the

above examples it is not necessary to use as assumed averages only those values which we have chosen for the purpose. Other values may also be so used The answers will not be different.

Advantages of Simple Arithmetic Average

1 It is easily understood and has a general usage

2 It is easy to calculate Its calculation is a common knowledge

3 It utilizes all the data in the group

4 It does not necessitate the arraying of data as the median does nor the grouping of data as the mode does

5 It can be known even when number of items and their aggregate values are known and details of the different items are not available

6 It is determinate It is not indefinite

7 The aggregate can be calculated if the number of items and the average are known

8 It affords a good standard of comparison since the abnormalities in opposite directions tend to cancel each other if the number of items is sufficiently large.

9 It is amenable to algebraic and arithmetic manipulation.

For all these qualities it is the most widely used average

Disadvantages of Simple Arithmetic Average

1 It may give considerable weight to extreme (and therefore abnormal) items A millionaire would greatly affect the average income of a town where a majority consists of ill paid artisans

2 It can hardly be located by inspection mode and median can be.

3 It can ignore any single item only at the risk of losing its accuracy Mode and median can be computed even when the values of extremes are not known

4. The average that results may not occur in the data at all, and may not therefore be representative to the fullest degree. The average of 2, 4 and 9 is 5, which does not occur in the series.

5. It cannot be used when the data are incommensurable: median can be used in qualitative studies.

6. This average might lead to fallacious conclusions when the actual figures from which it is obtained are not given. For instance, two students, A and B, get the following marks:—

		A	B
First Terminal Exam.	..	40%	60%
Second Terminal Exam.	..	50%	50%
Annual Examination	..	60%	40%

The average percentages of both of them are identical, 50. But A's progress is positive while B's is negative. If the average, 50, is not supported by the percentage marks in the three examinations, the fact that A is progressing while B is deteriorating would be concealed and a fallacious conclusion that the standard of both of them is the *same* would be drawn.

Uses of Simple Arithmetic Average.

It is used in many social and economic studies. Its use is daily routine in business and commerce. It is an average which even a 'man in the street' understands. It is the common average. Statistics uses it not only as a type for comparison, but for several other statistical calculations as well. "Average output of a commodity," "Average imports or exports over a period," "Average cost of production" "Average price"—in all such expressions the average used is the arithmetic average.

Weighted Arithmetic Average

In computing simple arithmetic average it was assumed that all items were of equal importance. This may not always be the case Where items vary in importance they must be assigned weights in proportion to their relative importance. The value of each item is then multiplied by its weight products summated and divided by the number of weights and *not* by the number of items The quotient is the weighted arithmetic average.

Symbolically

$$a' = \frac{\Sigma(wm)}{\Sigma\ (w)}$$

where a stands for weighted arithmetic average w for weight and m for measurement.

Weight is thus a number which stands for the relative importance of items. This relative importance may be real or estimated Accordingly weights are actual or approximate. Actual weights should be used where they are available otherwise they may be estimated on the strength of the best possible data available For instance if we know the *actual number of people engaged* on the teaching clerical and menial staff of an institution and the *average earnings* of each class of employees we should multiply the average earnings of each class by the actual employees in the corresponding class summate the products and divide the sum by the number of employees to secure the weighted arithmetic average of earnings The actual number of employees shall constitute actual weights The full method of working it out is shown in the following example.

Computation of Weighted Average

Example 13 Required to compute weighted arithmetic average from the data given in table 13 p 149 by the direct method.

Table 13. *Calculation of Weighted Average Earnings of the Employees of X College.*

1	2	3	4	5	6
Description of the employees	Number of employees w	Monthly Average Earnings m	Product of columns (2) and (3) wm	Estimated Weights w'	Product of columns (3) and (5) $w'm$
		Rs.	Rs.		Rs.
Professors ..	2	600	1,200	1	600
Lecturers ..	16	200	3,200	8	1,600
Demonstrators	4	100	400	2	200
Clerks ..	2	60	120	1	60
Peons ..	7	15	105	4	60
Watchmen ..	3	14	42	1	14
Totals ..	$\Sigma w=34$	$\Sigma m=989$	$\Sigma wm=5{,}067$	$\Sigma w'=17$	$\Sigma w'm=2{,}534$

Weighted arithmetic average—

(*a*) by using actual weights $= \dfrac{\Sigma(wm)}{\Sigma(w)}$

$= \text{Rs. } \dfrac{5067}{34}$

= Rs. 149-0-6.

(*b*) by using estimated weights $= \dfrac{\Sigma(w'm)}{\Sigma(w')}$

$= \text{Rs. } \dfrac{2534}{17}$

= Rs. 149-0-11.

Along with demonstrating the method of computing the weighted average, the above example also shows that it is not necessary that the weights applied should be actual ones (as

in column 2) they may be approximate also (as in column 5) the difference between the results obtained by using the actual and estimated weights being only five pies which is not material It should however be noted that when the number of weights used is small their size may have a considerable effect upon the average and therefore, if *estimated weights are used they should be approximately correct* If many weights are used the error in their estimation will be mostly unbiassed and, therefore cancelling one another The average would not then be materially affected

When we desire to calculate the average earnings per employee in example 13 above they might at first sight, appear to be $\frac{600+200+100+60+15+14}{6}$ or Rs 164.6 If it were so the total monthly earnings would be Rs ($164\ 6 \times 34$) or Rs 5596.4 but in fact the monthly pay roll amounts to (Rs. 600×2) + (Rs 200×16) + (Rs 100×4) + (Rs. 60×2) + (Rs. 15×7) + (Rs 14×3) that is Rs 5067 only Since Rs 164 6 multiplied by the number of employees does not yield the monthly pay roll it is not the correct average. If, however, we multiply the weighted average Rs 149-0-6 by the number of employees 34 we get Rs 5067 the monthly pay roll The weighted average of the average earnings in this case is approximately the same as the simple arithmetic average of the total earnings of all the 34 employees would be

In the above example we have multiplied the size of items (monthly average earnings) by their corresponding frequencies (number of employees) This has been called weighting by some writers[12] on statistics while it appears to be another device for computing the simple arithmetic average as told above A few writers do not regard it as weighting and their view is justified. Horace Secrist for instance is of opinion that weights should be 'determined by some evidence of importance other

[12] Notably, King, Boddington Connor

than that associated with the items themselves'.[13] Similarly, to Kelley weights are 'determined not at all, or not solely, by the population, but from other evidences of importance'.[14] Thus

Type of Employee	Number Employed	Relative Productivity	Productivity × Number
Male Adult ..	18	1	18
Female Adult ..	8	¾	6
Children ..	4	½	2

Therefore, men-equivalents = 26.

Similarly, a teacher may assign weights to different grades of work done by the students in proportion to the importance of the grades. He may, for example, assign 4 to seminar work, 2 to class-room work and 3 to monthly test. Marks obtained out of, say, 100 in each grade will be multiplied by the weight of the grade and the sum of the products divided by 9. This average will not correspond closely to the simple average as shown in example 14, p. 152.

The method used in example 13 for computing the weighted average may be termed the **direct method** (as distinguished from the **short-cut method)** which can be used to reduce the computations. In following the latter method any size of item may be assumed as the average, deviations of each item from this average computed and multiplied by the connected weight, the products summated and divided by the sum of weights, the quotient being added to the assumed average to arrive at the required weighted average.

[13] Horace Secrist, *An Introduction to Statistical Method*, New York, 1933, p. 280.

[14] Kelley, T. L., *Statistical Method*, New York, 1923, p. 68.

Symbolically,

$$a' - x' = \frac{\Sigma(wd_x')}{\Sigma(w)}$$

where, a stands for the weighted average, x', for assumed weighted average, w, for weight and d_x for deviation from assumed weighted average

Example 14 Required to compute weighted arithmetic average of the data given below by the short-cut method

Table 14. *Calculation of Weighted Average Marks of a Student*

Examination	Marks out of 100	Deviation from assumed average (50 marks)	Weight	Product of deviation and weight
	m	d_x'	w	wd_x'
Seminar	45	-5	4	-20
Class work	62	$+12$	2	$+24$
Monthly Test	52	$+2$	3	$+6$
	$\Sigma m = 159$		$\Sigma w = 9$	$\Sigma wdx' = 10$

Let the assumed weighted average be 50 marks Then

$$a' - x' = \frac{\Sigma(wd_x)}{\Sigma(w)}$$

$$\text{or} \quad a' - 50 = \frac{10}{9}$$

$a' = 51\%$ marks approx.

The simple arithmetic average in the above example will not approximate closely to the weighted average For,

$$a = \frac{\Sigma m}{n}$$

$$= \frac{159}{3}$$

$$= 53\% \text{ marks}$$

In fact, both the systems—allotting weights according to actual number and according to estimates of relative importance—are in vogue. We shall read more of them while discussing Index Numbers.

When Should Weighted Average Be Used?

(1) *When the items falling into different grades or classes of the same group show considerable variation, and it is desired to obtain an average representative of the whole group, weighted average is the only proper average to be used.*

It is usually found in a study of wages that the number of workmen earning high wage is much less than the number of those getting low wages. If, then, a simple arithmetic average of the wages in all the occupations—treating all grades as of equal importance—were computed, the wage of the manager would be given as much weight as that of a coolie or a gangman and the average would appear considerably large. Weights cannot be ignored in such cases.

(2) *When the means of the different grades constituting a group of observations and the number of observations falling in each grade are known, the only proper way to compute an average of the means is to strike the weighted average of the means.* Consider the following:—

I series: 2, 4. (Two items only).
II series: 5, 6, 7. (Three items only).

Simple average of I series $= \frac{2+4}{2} = 3$

Simple average of II series $= \frac{5+6+7}{3} = 6$

Simple average of the two series combined

$$= \frac{2+4+5+6+7}{5} = 4.8.$$

Simple average of the two averages $= \frac{3+6}{2} = 4.5$

Weighted average of averages of the two series

$$= \frac{(3\times2)+(6\times3)}{5} = 4.8$$

The simple average of the two averages, 4.5, no doubt gives an idea of the general average of the two series combined but the weighted average of averages, 4.8, is a better representative It also coincides with the simple average of the two series combined

The above principle also underlies example 13, solved above

(3) *Weighted average should also be used when the size of items changes and the relative proportion of the number of items also changes* For instance, if, in example 13, the number of employees in each grade were doubled or trebled so that the relative proportion of the number of employees remained unaffected, the average already computed, *viz*, Rs 149-0-6 would stand as representative even after the doubling or trebling of employees But, if their number changed such that their relative proportion were also changed, a new weighted average will have to be computed.

Caution in Weighting.

Weights should be used with caution since proper weighting is as valuable as wrong, manipulated or erroneous weighting is dangerous Weights should be as approximately accurate as possible and should be determined in the light of the best evidence available. We take below an example to demonstrate this principle

Example 15 Required to compute weighted arithmetic average of the percentage success in two universities

Table 15. *Calculation of Weighted Average of the Percentage Success in X and Y Universities.*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
University Examination	Relative proportion of candidates	Percentage of success in X university	Product of columns 2 and 3	Percentage of success in Y university	Product of columns 2 and 5	Arbitrary weights	Product of columns 3 and 7	Product of columns 5 and 7
M.A.	10	80	800	75	750	15	1200	1125
M.Sc.	7	65	455	50	350	10	650	500
B.A.	40	60	2400	70	2800	10	600	700
B.Sc.	25	55	1375	75	1875	5	275	375
B.Com.	13	75	975	65	845	40	3000	2600
Totals	95	335	6005	335	6620	80	5725	5300

I Simple average of percentage success in—

X University $= \frac{335}{5} = 67$

Y University $= \frac{335}{5} = 67$

II Weighted average, by using the weights in columns 2, in—

X University $= \frac{6005}{95} = 63.2$

Y University $= \frac{6620}{95} = 69.7$

III Weighted average, by using the weights in column 7, in—

X University $= \frac{5725}{80} = 71.6$

Y University $= \frac{5300}{80} = 66.3$

The above example makes the following facts clear:—

(i) The simple averages for both the universities are identical. If they are used for comparing the

percentage success it would appear that the average percentage success in both of them is the same

(ii) But when weights are assigned to the results in proportion of the number of candidates at different examinations Y university appears to have much better result than X This conclusion is corroborated by the fact that in the B.A. and B Sc examinations where the number of candidates is very large the percentage success in Y university is higher—much higher in B Sc.—than in X These weighted averages therefore yield a proper comparison.

(iii) When weights given in column 7, chosen quite arbitrarily without regard to the proportionate importance of the items are used the conclusion arrived at in (ii) above is reversed X university appears to indicate better results than Y This inference is not corroborated by facts Therefore allotting of weights needs great care and caution lest fallacious conclusions result

(iv) A comparison of the averages arrived at by using the weights given in columns (2) and (7) reveals that the averages for X university in II and III cases 63 2 and 71 6 and also the averages for Y in the two cases 69 7 and 66.3 are very much different between themselves It is so because the weights in column (7) do not have between one another the same proportion as weights in column (2) do In our example of the earnings of the employees of X college table 13 page 149 the weights in column (5) have the same proportion among themselves as those in column (2) do. The averages resulting from the use of the two weights are therefore not materially different It is thus

established that **it is not the absolute size of the weights but their relative size that is important. A weighted average is unaltered if all the weights are multiplied or divided by the same quantity,** that is, if their mutual proportions remain unchanged. Even if the multiplication or division is not accurate but only approximate as in column (5), table 13, the resulting weighted average shall be approximately correct.

Difference between Simple and Weighted Averages.

Strictly speaking the so-called "unweighted" (simple) average is a special case of "weighted" average in which the weights are unity for each item. In the "weighted" average, weights are chosen according to some systematic plan and are not unity for each item.

However, in example 13, page 149, the simple average was Rs. 164.6 while the weighted average was Rs. 149-0-6. Similarly, in the example considered on pages 153-154 the simple average of 3 and 6 was 4.5 whereas their weighted average was 4.8. It is thus established that any weighted mean will generally differ from the unweighted average of the same quantities This difference, in general terms, may be expressed as follows:

If the weights are so distributed that the larger items are associated with smaller weights, and the smaller items with larger weights, the smaller items will have greater influence in the formation of the weighted average and the weighted average will be less than the simple average. If, on the other hand, larger items have larger weights and the smaller ones have smaller weights, the larger items will have more influence and the weighted average will be more than the simple average. But if there is apparently no connection between weights and items, that is, if larger weights are as much

associated with larger as with smaller items and similar is the case with smaller weights, the difference between the weighted and the simple averages may be insignificant.

Thus the simple average of five items 1, 2, 3, 4 and 5 is 3; their weighted average if their weights be respectively 5, 4, 3, 2 and 1 is 2.33 which is less than the simple average; their weighted average if their weights be respectively 1, 2, 3, 4 and 5 is 3.66 which is more than the simple average; and their weighted average if their weights be respectively 3, 5, 1, 2 and 4 is 2.93 which is not materially different from the simple average.

This appreciation of the difference between the simple and weighted averages is of great value in certain cases. For instance, if we have reason to believe that weights are not connected with the size of items in a particular way, that is, as likely as not large and small weights are attached indiscriminately to large and small sizes of items, we shall not be unjustified in taking the simple average as a fairly good approximation to the unknown weighted average. We shall then avoid the strain involved in discovering exact weights.

Characteristics and Uses of Weighted Average.

The weighted average, generally speaking, possesses the same advantages and disadvantages and has almost the same uses as the simple arithmetic average. Like the latter it gives the central tendency of the whole group to which it pertains and does not refer to the actual conditions of any one grade or class of a group. It is useful for general comparison. It is a good average to use when a group as a whole, say the whole of an industry, is surveyed. If, however, we wish to study the actual condition of the various grades or classes of a group, we should compute the averages of the different grades separately and compare such averages. That is, we should study each homogeneous part separately.

The weighted average is invariably applied in the calculation of birth, marriage and death rates, and their comparison in different places or at different times. Its application to death rates will be treated towards the end of this chapter, and to index numbers in Chapter XII.

It may be useful to remember that weighting is essential for attaining accuracy in the result when the series is small; it is unnecessary in a very long series since in such series the weighted and simple averages tend to be identical.

THE GEOMETRIC AVERAGE

The geometric average, also called the geometric mean, is the nth root of the product of the n quantities of a series. The geometric mean is obtained by multiplying the values of the items together and extracting the root of the product corresponding to the number of items. Thus, the square root of the product of two quantities is their geometric mean. Similarly, the cube root of the product of three quantities is the geometric mean of three quantities. Symbolically,

$$g = \sqrt[n]{a \times b \times c \times \ldots\ldots\ldots n}$$

where, g stands for the geometric mean; n, for the number of items and a, b, $c \ldots\ldots n$, for the values of n items.

The geometric mean of 4 and 9 is equal to $\sqrt{4 \times 9} = 6$; the geometric mean of 2, 4 and 8 is equal to $\sqrt{2 \times 4 \times 8} = 4$. When the number of items in a series is larger than three, this process is difficult to follow. To obviate the difficulty, logarithm of each size is obtained from a Mathematical Table.[15] The logarithms of all the values are added up and divided by the number of items. The anti-logarithm of the quotient is the required geometric mean. The formula is:

$$g = \text{Anti-log}\left[\frac{(\log a + \log b + \log c + \ldots.\log n)}{n}\right]$$

[15] Mathematical Tables are given at the end of the book.

Example 16 Required to calculate the geometric average of the data given below in the first column

Table 16 *Calculation of Geometric Average*

Size of Items	Logarithms
4 5	6532
250.0	2 3979
12 0	1 0792
119 5	2 0792
30.0	1.4771
42 0	1 6232
75 0	1 8751
35.4	5490
Rs 568.4	12 7339

According to the formula we have

$$g = \text{Anti log} \frac{12\ 7339}{8},$$
$$= \text{Anti log } 1\ 6$$
$$= \text{Rs } 39\ 81$$

The geometric mean is always less than the simple arithmetic average unless all the sizes of the variable are equal in magnitude Thus in the above example

$$a = \text{Rs } \frac{568.4}{8} = \text{Rs } 71\ 05$$

which is greater than the geometric mean

Weighted Geometric Mean

To compute the weighted geometric mean of a series of items each individual item should first be multiplied by its corresponding weight and the products obtained should be multiplied by one another The *n*th root of this final product where *n* stands for the total number of weights is the required weighted geometric average. Symbolically

$$g' = \sqrt[n]{an_1 \times bn_2 \times cw_3 \qquad nw_n}$$

where, g' represents weighted geometric mean, $w_1, w_2 \ldots\ldots w_n$ represent the weights corresponding to the size of item to which they relate, n represents the total number of weights.

In practice, logarithms may be used. First, logarithm of each individual item should be found from a mathematical table. Each log. should then be multiplied by its weight. The summation of such products divided by the total number of weights is the required weighted geometric mean. This may be expressed as follows:

$$g' = \text{Anti-log.} \left[\frac{\log a \times w_1) + (\log b \times w_2) + \ldots\ldots (\log n \times w_1)}{w_1 + w_2 + \ldots\ldots w_n}\right]$$

Example 17. Required to compute weighted geometric mean of the data given below.

Table 17. *Calculation of "Capital" Index of Indian Industrial Activity, March* 1942, *Based on Geometric Mean.*

Items	Weights	Index No. (1935=100)	Log.	Weight × Log.
Indian Cotton Consumption ..	9	149.5	2.1761	19.5849
Jute manufactures ..	6	134.9	2.1303	12.7818
Steel Ingots ..	5	147.1	2.1673	10.8365
Pig Iron .	8	134.4	2.1271	17.0168
Paper ..	3	185.2	2.2672	6.8016
Coal ..	7	110.0	2.0414	14.2898
Rail & River borne Trade ..	24	108.9	2.0374	48.8976
Cheque Clearances ..	20	88.5	1.9469	38.9380
Notes in Circulation*	6	132.4	2.1206	12.7236
Consumption of Electricity ..	7	152.9	2.1847	15.2929
	95			197.1635

* April, 1935 to March, 1936=100.

According to the formula we have

$$g' = \text{Anti log} \frac{197\ 1635}{95}$$
$$= \text{Anti log}\ 2\ 0754$$
$$= 118\ 9$$

The Capital Index of Indian Industrial Activity for March 1942 therefore is 118 9

Advantages of the Geometric Mean.

1 It is determinate, provided the values of the variable are greater than zero

2 It is based on all the data in the group

3 It gives less weight to large items and more to small ones than does the arithmetic average

4 It is particularly useful when dealing with ratios

5 It is amenable to arithmetic and algebraic manipulation.

Disadvantages of the Geometric Mean.

1 It cannot be used when any of the quantities is zero or negative for when a quantity is zero the product of all quantities will be zero and the g will be zero and when a quantity is negative the product of all quantities will be negative and the g will become unrepresentative and imaginary

2 It may be found to lie at a point where very few (or even none) of the actual measurements lie

3 It entails much work of calculation and is difficult of computation

4 It is less easily understood than the arithmetic average

Uses of the Geometric Mean.

The property of giving large weight to small items makes geometric average a very suitable type in studying various social and economic phenomena where it is desired to give large weight to small items If some items in a series are

very big and others very small it is not the arithmetic average, median or mode but the geometric mean that yields a representative type. If the annual incomes of, say, the employees of a university vary between Rs. 180 and Rs. 24, 000, geometric mean of the incomes will give a good idea of their average yearly income. If arithmetic average were used, a single salary of Rs. 24,000 would pull the arithmetic mean very high, because of the comparatively very low salaries of clerks and peons. The geometric mean would nullify the effect which large values have upon the arithmetic average. It may be remembered that if in a series the arithmetic and the geometric means are found to differ considerably from each other, the geometric mean should be regarded as better representative of the two, since it falls within a range of the majority of the given examples.

Another important use of geometric mean is in connection with index numbers. Index numbers are ratios, and the geometric mean is particularly useful in dealing with relative as against absolute differences.

Geometric mean is used in the construction of the Index of Indian Industrial Activity by the *Capital* and in that of the "Wholesale Price Index Numbers of certain articles in India" by the Economic Adviser to the Government of India. It is used in the *Board of Trade* Index Number of Wholesale Prices in Great Britain. It was used by Professor W. S. Jevons in his study of the changes in the general level of prices. The difficulty experienced in its calculation and the fact that it is too abstract to be readily intelligible have stood in the way of its popularity and general use. It is, however, useful in the averaging of ratios and rates of interest.

THE HARMONIC AVERAGE

The Harmonic Average, also called the harmonic mean, is the total number of items of a variable divided

by the sum of the reciprocals of the values of the variable Symbolically,

$$h=\frac{n}{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\qquad\frac{1}{n}}$$

where, h stands for the harmonic mean, a, b, c n represent the values of the n items of the variable, and n is the number of items Reciprocals of numbers can be easily obtained from a Mathematical Table [19]

The harmonic mean can also be expressed as the reciprocal of the arithmetic average of the reciprocals of the values of the items of a series Thus

$$h=\text{Reciprocal}\frac{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\dots\quad\dots\frac{1}{n}}{n}$$

Example 18 Required to calculate the harmonic mean of the data given in table 16, page 160.

Table 18 *Computation of Harmonic Mean*

Sizes of Items Rs	Reciprocals.
4.5	.2222
250.0	.0040
12.0	.0833
119.5	.0084
30.0	.0333
42.0	.0238
75.0	.0133
35.4	.0283
	.4166

[19] The Mathematical Tables given at the end of the book give reciprocals of natural numbers

According to the formula we have,

$$h = \frac{8}{\cdot 4166} = Rs.\ 19\cdot 2$$

$$\text{or}\ \ h = Reciprocal\ \frac{\cdot 4166}{8}$$

$= $ Reciprocal .05208
$= $ Rs. 19.2.

The harmonic mean is always less than the geometric mean. In the above example:

$a =$ Rs. 71.05
$e =$ Rs. 39.81
$h =$ Rs. 19.2

Characteristics and Uses of Harmonic Mean.

The harmonic mean is determinate and considers the values of all items of the data. It gives the largest weight to the smallest items, and is valuable where such weighting is desirable. It may be used in averaging of rates and time. It is used in very special cases and is not suitable for general application. The time and trouble involved in its calculation also stand in the way of its popularity. It is abstract and not easy to understand. It may not be an actual example occurring in the series.

Averages of the First Order.

The various averages discussed so far are averages of the "first" order—that is, they deal with the actual values of a statistical variable. In contrast to them we shall later (in Chapter XI) study averages of the "second" order—that is, those which summarize not the actual values but the difference between them and some average. Averages of the "first" order can be used as representatives or substitutes of the data to which they relate.

Typical and Descriptive Averages.

It is always arithmetically possible to calculate an average from a given series But this does not imply that every average is statistically significant If the average of a series is found to lie near a point round which the data exhibit a tendency to cluster the average may be presumed to be sufficiently representative of the series It is then called a Typical Average If on the other hand the distribution of items is irregular so that the data seem to cluster round several points or do not cluster at all the average has only arithmetical significance and should not be considered as full representative of the series It is then called a Descriptive Average. Typical average can be substituted for the series for purposes of comparison or for other information relating to the series.

Relative Position of Averages.

As shown in example 18, above the arithmetic average is greater than the geometric mean of the same quantities and the geometric mean is less than their harmonic mean These three means will be identical if all the sizes of items in the given series are of equal value Symbolically this may be expressed by the inequalities

$$a \geqq g \geqq h,$$

where the sign of equality prevails only when all the sizes are equal

If a given group is symmetrical the arithmetic mean mode and median are coincident In a moderately asymmetrical distribution the following relationship between the mean mode and median may be found to hold good

$$(M-Z) = \tfrac{2}{3}(a-Z)$$

If there is a small group at a high degree the arithmetic mean would probably be above the median and the mode below

it. The arithmetic average would probably be below the median and the mode above it if there is an absence of high numbers. It should be clearly noted that these rules are only tentative.

Choice of Average.

An average is a simple comprehensive expression of a series of divergent individual values. All averages do not characterize the series in the same way. They yield only that information which, by their nature, they are able to transmit. This information differs according to the kind of average used. Therefore, it is the *purpose* for which an average is to be employed that will largely determine the choice of an average. No one average is good for all purposes. The mean chosen to study a given phenomenon should be such as will throw light on the striking features and the essential characteristics of the phenomenon. Different averages will apply to different groups of data; each must be selected on its own merits. Each average is affected differently by the distribution, frequencies and the character of the details. A knowledge of its peculiarities or characteristics is, therefore, a pre-requisite for scientific use of an average. It is evident that an average simplifies complexity, but if the particular merits, demerits, scope and characteristics of the average are neglected, the simplicity arrived at shall not be worth having. Caution, foresight and analysis are, therefore, necessary in the use of averages. If they are ignored, the very principles on which scientific method rests shall be violated. This is not desirable.

What are the desirable properties for an average to possess? First, it should be rigidly defined; second, it should be based on all the observations of the data; third, it should be readily comprehensible; fourth, it should be capable of being computed with reasonable ease and rapidly; fifth, it should be as little affected by fluctuations of sampling as possible; and last, it should be readily amenable to arithmetic or algebraic treatment.

From a perusal of the advantages and disadvantages of the various averages outlined in the foregoing pages it will be evident that the arithmetic average—the common mean—possesses the above properties more than any other single average does It is rigidly defined is based on all observations is readily comprehensible is less affected by fluctuations of sampling than say the median and above all is suitable for algebraic treatment Of course median is somewhat more easily computed than the arithmetic average but median is often indeterminate and its algebraic treatment is difficult if not impossible in many cases Mode is hardly useful in elementary work owing to the difficult of locating it with precision Since the arithmetic average uses all the items of a given series it is likely to be less erratic i e less sensitive to small change in values of individual items The arithmetic average is therefore quite suitable for all general purposes unless there is special reason to select any other average For instance if items of small values are far larger in a series than items of large values arithmetic average will not be a good average to use Instead the geometric mean will be used And if it is necessary to give more weight to the smallest item than to other ones harmonic mean will be the proper average to compute. Similarly if enquiry is made into the average size of shoe sold at a shop or an 'average coat tailored at a tailor's it is not arithmetic average but mode that will serve the purpose Again if an idea of average intelligence of a class is to be had median shall be the best average since it can be used even in those cases where the data are not quantitatively measurable These are typical cases where arithmetic average should not for special reasons be used In general arithmetic average is suitable for most purposes

Limitations of Averages

It is evident that an average is a summary of the details of a series It is used as a substitute for what it replaces

But here lies its limitation. Different details may yield the same average, yet it is the details which may be of interest. An average, if at all it does, rarely contains as much significance as the individual items do. If averages are used alone, not supported by the details, since they are merged in the single simple expression, are ignored except in so far as they are reflected in the summary. Averages, therefore, do not relate the whole story. They indicate only the central position of a group. What lies behind them is not their task to reveal.

It follows that in computing and using an average one should know the following things if one is to guard himself against a fallacy of argument:

(1) The purpose of the average.
(2) The peculiarities of the data to be summarized.
(3) The characteristics of each average.
(4) A deep knowledge of the whole subject to which the given data relate in order to be certain that the average computed shall be significant and suitable.
(5) The extent to which data are homogeneous.

Standardized Death Rate.

If death rate is calculated for each age-group of a locality's population, and then death rate is calculated for the whole of the population by the use of weighted average, the latter death rate is called **General Death Rate** or **Crude Death Rate.** If this crude death rate for a locality is compared with that for the standard population (*e.g.* the population of the country at large, or, of another locality assumed to be standard), misleading conclusions might result. To avoid fallacious comparison it is advisable to eliminate differences between age compositions of the populations of the two localities by applying the local death rates in each age group to the standard population. The following is a simple illustration—

Example 19 Required to compute crude and corrected death rates

Table 19 *Computation of General and Standardized Death rates*

Age-group Years	Standard Population A: Population	Deaths	Death rate per 1000	Local Population B: Population	Deaths	Death rate per 1000
Under 5	600	18	30	400	16	40
5—15	1000	5	5	1500	6	4
15—65	3000	24	8	2400	24	10
Above 65	400	20	50	700	21	30
Total	5000	67	13.4	5000	67	13.4

General Death Rate of Standard Population

$$=\frac{1}{5000}\ (600\times30+1000\times5+3000\times8+400\times50)$$
$$=13.4 \text{ per } 1{,}000$$

General Death Rate of Local Population

$$=\frac{1}{5000}\ (400\times40+1500\times4+2400\times10+700\times30)$$
$$=13.4 \text{ per } 1{,}000$$

Upon comparison of the two general death rates, computed by using the weighted average, nothing remarkable will be noted both the populations have the same death rate And, if death rate is any measure of the *health* of a population both the populations are equally healthy To justify this viewpoint one might add that the total number of inhabitants in both the places, 5000 is the same that the total number of deaths in the two cases, 67 is identical and that the number

of deaths in the age-group (15—65) years is equal in both of them. With these arguments one might try to make others believe that both A and B are equally healthy. But it should be noted that the death rates in different age-groups in both the places are different and also that the distribution of population in the two places into various age-groups is not identical. That is, the basis of comparison is not the same. It is, therefore, not fully correct to believe that both the towns are equally healthy unless this conclusion is found to hold good when the basis of comparison is made identical. To do so, we eliminate the differences between age constitutions by assuming that the distribution of local population into different age-groups is the same as that of the standard population. Then, by applying the local death rates to the changed distribution we calculate another weighted average death rate, now called the **Corrected** or **Standardized Death Rate.** Thus,

Standardized Death Rate of Local Population

$$= \frac{1}{5000}(600\times40+1000\times4+3000\times10+400\times30)$$

$$=14 \text{ per thousand.}$$

The standardized death rate of local population is higher than the crude death rate of standard population, leading us to conclude that the local population is less healthy than the standard population.

Similarly, there could be a case where the general death rates of A and B were different, but the death rates for different age constitutions were the same. This paradox, again, would have been due to differences in the distribution of population into the various age-groups. The paradox would be removed by computing the standardized or corrected death rate of the local population.

This method is of general application. We have applied it to death rates. We may standardize marriage rates or unemployment rates as well.

EXERCISES

(1) What is an average? How does it differ from a percentage? What purposes does it serve?

(2) What do you understand by homogeneity of data? Should the data from which averages are computed be homogeneous? Give reasons

(3) Define Mode, Median Mean Geometric average and Harmonic mean and clearly explain their uses.

In which problems can each one of them be used with the greatest advantage?

(4) Compare the advantages and disadvantages of the different averages

(5) What is a statistical average? What are the desirable properties for an average to possess? Which of the averages, you know possesses most of these properties?

(B Com Alld, 1944)

(6) How will you locate the mode when the distribution of frequencies for class intervals whose magnitude is one inch gives three maxima? Take an hypothetical example to explain the whole process

(7) How will you locate the median when
 (a) the number of items in a series is even
 (b) the series is a discrete one,
 (c) only the frequency distribution of a series is given?

(8) What are Quartiles Deciles and Percentiles? What information do they give regarding a series? How are they calculated in (i) series of individual observations (ii) discrete series and (iii) continuous series Show their relationship with the median

(9) How will you compute the simple arithmetic average of
 (i) a series of individual observations
 (ii) a discrete series and
 (iii) a continuous series?
Explain the direct as well as the short cut methods

(10) Define weighted average and explain how it differs from simple mean. Give the method of its computation and point out the cases in which weighted average should be used.

(11) Differentiate between Crude (General) and Corrected (Standard) death rates.

How is the principle of weighting applied to the determination of standardized death rates from crude death rates?

(12) Discuss critically the use of weighted mean in statistics. (B.Com, Cal., 1937).

(13) Explain clearly the meaning, the best method of computing, and the use of each of the following —

(*i*) the arithmetic average.
(*ii*) the median, and
(*iii*) the mode.

(B.Com., Agra, 1934).

(14) State the formulæ of the principal forms of averages employed in Statistics, and explain, so far as you can, the principles upon which they are based.

(15) What are the limitations of the uses of each one of the different kinds of average?

(B.Com., Alld., 1939).

(16) Which average would you use in studying the following problems and why?

(*a*) Comparing the economic condition of India with U. K.'s.
(*b*) Size (number of members) of an average family.
(*c*) Size of average agricultural holding.
(*d*) Average marks in an examination.
(*e*) Average height or weight of students.
(*f*) Average length of the leaves of a tree.
(*g*) Average intelligence.
(*h*) Average sales of a shopkeeper.

(17) Calculate the median, arithmetic average and quartiles of the series relating to heights of 53 students given in exercise 17, Chapter VIII, page 108.

(18) What is the purpose served by an 'average'? Discuss the special advantages attached to the different averages, and illustrate their uses.

(B.Com., Agra, 1942).

(19) Calculate the geometric and arithmetic means of the series given in exercise 1, Chapter VI, page 65

(20) According to the census of 1941 following are the population figures in thousands, of first 36 cities of India

2488	591	437	208	213	143
1490	407	284	176	,169	181
777	387	302	213	204	153
733	301	263	176	178	142
522	360	260	193	131	92
672	258	239	160	147	151

Find the median and quartiles

(21) Find the geometric and harmonic means of the series given in exercise 20, above

(22) Compute the mode and arithmetic average of the following series Account for their difference

Size of item	Frequency
2	3
3	8
4	10
5	12
6	16
7	14
8	10
9	8
10	17
11	5
12	4
13	1

(23) The following table gives the marks obtained by a batch of 25 students in a certain class test in Economics and Politics —

Roll Number of the Students	Economics	Politics
1	29	36
2	65	30
3	33	38
4	45	39
5	51	64
6	72	50
7	49	46
8	33	15

Roll Number of the Students	Economics	Politics
9	42	42
10	25	10
11	28	72
12	35	33
13	46	80
14	47	44
15	60	85
16	30	20
17	32	32
18	52	25
19	54	55
20	56	28
21	58	53
22	49	35
23	38	40
24	40	62
25	46	58

In which subject is the level of knowledge of the students, as revealed from the above figures, higher? Give reasons.

(M.A., Alld., 1937).

(24) Find out the Mode of the following series :—

Size	Frequency	Size	Frequency
5	48	13	52
6	52	14	41
7	56	15	57
8	60	16	63
9	63	17	52
10	57	18	48
11	55	19	40
12	50	..	..

(B.Com , Alld , 1943).

(25) Compute the weighted geometric average of Relative Prices of the following commodities for the year 1939 (Base year 1938 —Price 100) :—

Commodity	Relative Price	Weight (value produced in 1938)
Corn	128.8	1,385
Cotton	62.4	819
Hay	117.7	842
Wheat	99.0	561
Oats	130.9	408
Potatoes	143.5	194
Sugar	125.6	142
Barley	150.2	100
Tobacco	101.1	103
Rye	116.2	25
Rice	117.5	17
Oil Seed	78.7	29

How does it differ from the unweighted geometric mean and why?

(B Com Alld. 1943)

(26) Statistics help collecting agreements of wage adjustments What data are required for the consideration of a revision in wage rates in a factory which average will you utilize and why?

(M A Agra 1940) M Com., Alld, 1943)

(27) Compare the relative advantages and disadvantages of the Arithmetic mean the Median and the Mode

The following table gives the results of certain examinations of three universities in the year 1936. Which is the best university? Give reasons for your answer

University Examination	Percentage results in the University		
	A	B	C
M A	80	75	70
M Sc	70	70	60
B A.	65	80	70
B Sc.	60	70	80
B Com.	75	65	75

(M.A. Cal 1937)

(28) The following table gives the number of persons with different incomes in the U.S.A. during the year 1929.

Income in thousands of dollars	No. of persons in Lakhs.
Under 1	13
1 — 2	90
2 — 3	81
3 — 5	117
5 — 10	66
10 — 25	27
25 — 50	6
50 — 100	2
100 —1000	2

Calculate the average income per head.

(B.Com., Luck., 1939).

(29) The marks obtained by students of classes A and B are given below. Give as much information as you can regarding the composition of the classes in respect of intelligence:—

Marks obtained	No. of students in class A	No. of students in class B.
5 — 10	1	5
10 — 15	10	6
15 — 20	20	15
20 — 25	8	10
25 — 30	6	5
30 — 35	3	4
35 — 40	1	2
40 — 45	0	2

(B.Com., Agra, 1939).

(30) Explain what is meant by *weighted average*, and discuss the effect of weighting.

Calculate (i) the *unweighted mean* of the prices in column III and (ii) the mean obtained by weighting each price by the quantity consumed and explain why they differ as they do —

I Articles of Food	II Quantity Consumed	III Price in Rupees per maund
Flour	11 5 mds	58
Ghee	5 6 mds	58.4
Sugar	·28 mds	8.2
Potato	·6 mds	2 5
Oil	35 mds.	20 0

(M.A., Cal. 1937)

(31) Explain the short-cut method of calculating the arithmetic average

The following data relate to sizes of shoes sold at a store during a given week. Find the average size by the short-cut method

Size of Shoes	4·5	5	5·5	6	6·5	7	7·5	8	8·5	9	9·5	10	10·5	11
No of pairs	1	2	4	6	15	30	60	95	82	75	44	25	15	4

(M.A. Cal., 1936)

(32) The following table gives the population of males at different age-groups of the U K. and India at the time of the Census of 1931

Age-Group	U K. Lakhs	India Lakhs
0 — 5	18	214
5 — 10	19	258
10 — 15	20	222
15 — 20	18	157
20 — 25	16	145
25 — 30	14	161
30 — 40	27	257
40 — 50	25	184
50 — 60	19	120
Above 60	17	100

Compare the average age of males in the two countries and account for the difference if any

(B Com. Alld 1936)
(B Com. Luck 1941)

(33) From the following table calculate the average price of a lb. of biscuits and also the weighted average price.

Price per lb. Rs. a. p.	lbs. sold
0—10—0	100
0—15—0	87
1— 2—0	63
1— 4—0	59
1— 8—0	49
2— 0—0	19

Which of the two averages gives a better indication of the average price?

(34) Following are the lengths in inches of 102 *nim* leaves. Calculate the median, 6th decile and 37th percentile.

1.85, 1.5, 1.95, 1.9, 1.6, 2.2, 2.45, 2.72, 2.48, 3.0, 3.7, 3.0, 2.85, 3.25, 2.48, 3.43, 2.80, 2.35, 2.64, 2.76, 2.9, 2.6, 2.65, 2.95, 2.70, 2.50, 1.95, 1.95, 1.58, 2.45, 2.92, 2.95, 2.78, 2.60, 2.54, 2.79, 2.90, 3.05, 2.82, 2.33, 2.90, 2.88, 2.15, 1.75, 2.40, 2.48, 2.15, 2.65, 2.50, 2.20, 2.40, 2.45, 2.5, 2.56, 2.40, 2.25, 2.30, 1.50, 1.90, 2.30, 2.88, 2.30, 1.95, 1.85, 2.95, 2.90, 2.00, 2.80, 3.25, 2.95, 3.20, 2.85, 2.70, 2.77, 2.44, 2.10, 2.54, 2.70, 2.40, 2 65, 2.60, 2.94, 2.05, 2.06, 2.50, 2.30, 1.90, 2.78, 2.60, 2.35, 2.72, 2.85, 2.70, 2.50, 2.53, 1.98, 2.94, 3.05, 2.66, 2.85, 3.10.

(35) Calculate the arithmetic average of the above series by (*i*) the direct method, and (*ii*) the short-cut method.

(36) Calculate the arithmetic average and mode of the following series.

Size of item	Frequency
3— 5	14
— 7	16
7— 9	25
9—11	22
11—13	12

(37) Compute the geometric and harmonic means of the following items

3755 153 285 12.01 45 37, 1279 35 419 53

(38) If in ten successive years the quantities 1, 2 3 4 5 6 7 8 9 10 are sold at prices 1 2 3 4 5 6 7, 8 9 10 what are the weighted average and simple arithmetic average prices?

(39) Find the mode and arithmetic average of the daily wages in the following series

Daily Wages in Annas	Frequency (Number of employees)
3	2
5	10
7	12
9	15
11	20
13	13
15	12
17	10
19	4

(40) Show the relative positions of different averages in moderately symmetrical series

Find the mode median and arithmetic average of the following series and state if the series is symmetrical

Size of item	Frequency
10	30
11	35
12	38
13	42
14	46
15	42
16	38
17	35
18	30

(41) Which of the two places for which the mortality data are given below would you describe as more healthy than the other? Give reasons.

Age in Years	Town X (Standard)		Town Y (Local)	
	Population	Deaths	Population	Deaths
under 10	15,000	375	10,000	300
10—30	50,000	250	52,000	312
30—70	120,000	840	126,000	1,008
above 70	15,000	975	12,000	840

(42) Compute the Crude and Standardized death-rates in the following, and state if local population has higher or lower death-rate.

Age group (years)	Standard Population		Local Population	
	Population	Deaths	Population	Deaths
under 5	6,000	150	2,500	63
5—15	10,000	20	12,500	25
15—65	12,500	50	20,000	80
above 65	4,000	160	5,000	200

(43) *Value of Exports & Imports of Commodities for India for 1934-35.*

Months	Exports Crores of Rs.	Imports Crores of Rs.
April	.. 12.7	10.9
May	.. 13.3	10.5
June	.. 12.5	9.5
July	.. 12.8	9.9
August	.. 12.3	10.7
September	.. 12.1	10.5
October	.. 12.4	12.5
November	.. 12.3	11.4
December	.. 12.2	10.3
January	.. 13.7	12.9
February	.. 13.2	10.6
March	.. 15.6	12.4
Total	.. 155.1	132.2

Calculate the median and arithmetic average of the above figures of exports and imports separately

(44) The following are the weekly market values of the shares of Imperial Bank of India (paid up value Rs 500) from Jan 4 1933 to Dec. 20 1933

1 35 1235 1236 5 1261 5 1266 5 1166 1190 1176 1160, 1186 1221 5 1220 1234 1230 1235 1236 5 1232 5 1251 5 1244 1231 5 1216 5 1216 5 1221 5 1221 5 1205 1184 1196 1212.5, 1207 1192 1195 1206 1201 1196 1198 1208.5 1202 5 1212 5, 1234 5 1225 1230 1220 1220 5 1220.5 1240 1252 5 1246, 1246 5 1246 5 1230.5

(from the Capital)

Calculate the arithmetic mean and geometric average of the above series

(45) The following table gives the age distribution of widows in India (Census Report 1931) Calculate the median age of the widows and also the upper and lower quartiles

years	No of widows
0—10	135,862
10—20	718 101
20—30	2,456,835
30—40	4 847 631
40—50	6,480,259
50—60	5 908 159
60—70	3 743 615
70 & over	1,957 506
Total	26,247 968

6 (46) The following table gives the marks obtained by a batch of 30 B Com. students in a class test in Statistics (Marks 100)

Roll No.	Marks obtained	Roll No.	Marks obtained
1	33	16	24
2	32	17	33
3	55	18	42
4	47	19	38
5	21	20	45
6	50	21	26
7	27	22	33
8	12	23	44
9	68	24	48
10	49	25	52
11	40	26	30
12	17	27	58
13	44	28	37
14	48	29	38
15	62	30	35

Find the values of the Mode, the Median, and the Quartiles,

(B.Com., Alld., 1938).

(47) Calculate the arithmetic average and the geometric mean of the series given in exercise 46 above. Which average represents the series better?

(B.Com., Alld., 1938).

(48) The following table gives the distribution of population according to age in India and Japan at the time of the last census (1931):

Age group in years			Population in millions in India	Japan
0—10	..	..	98.9	17.8
10—20	..	..	72.5	14.3
20—30	..	..	63.2	11.3
30—40	..	.	48.6	8.6
40—50	..	...	32.6	6.5
50—60	..	..	19.4	5.4
60—80	..	..	13.2	5.1

Calculate the average age of people in India and in Japan, and comment on the difference.

(B.Com., Alld., 1940).

(49) In order to decide whether one city is more healthy than another on the basis of death-rates what information would

you require in addition to the total number of deaths and the total population of the two cities? How would you use this information to decide which city is more healthy?

(M A Alld 1935)

(50) Calculate the simple average and the weighted average of the following items —

Item	68	85	101	102	108	110	112	113	124	128	143	146	151	153	172
Weight	1	46	31	1	11	7	23	17	9	14	2	4	6	5	2

Account for the difference in the two averages

(M A Alld., 1940)

(51) The following is the distribution of wages per thousand employees in a certain factory

Daily wages in Annas	2	4	6-	8	10-	12	14-	16-	18-	20-	22	24-	Total
Number of employees	3	13	43	102	175	220	204	139	69	75	6	1	1000

Calculate the modal and median wages and explain why there is a difference between the two

(M A Alld 1940)

(52) Calculate the median 3rd quartile 4th decile and 80th percentile of the following series

Size of item	Frequency
6—10	20
11—15	30
16—20	40
21—25	40
26—30	10

(53) Calculate the average median and upper and lower quartile ages in the following table

Age-groups	Population in thousands 1881	1931
0—4	3520	3280
5—9	3160	3500
10—19	5340	7200
20—29	4560	6640
30—39	3420	5980
40—49	2660	5240
50—59	1900	3780
60—69	1320	2440
70—79	600	1220
80 & over	120	320

(M.A., Agra, 1940).

(54) Point out the ambiguity or mistake, if any, in the following statements:

(*i*) There are 250 employees in a sugar mill. Their daily earnings are about Re. 1/ per man on an average. Therefore, their total monthly earnings are Rs. 7,500.

(*ii*) An ordinary person in India consumes one chhatak of pulse per day. Therefore, the total quantity of pulse consumed by India's 40 crores of people every year is about 23 crores of maunds.

(*iii*) The monthly expenditure of the vast majority of students in a university is Rs. 50. Therefore, the total monthly expenditure of 2,000 students is Rs. 1,00,000.

(*iv*) A cloth dealer usually receives 150 customers a day. Therefore, the total number of customers he receives in a month is 4,500.

(55) The following table gives the value of imports of commodities into India in crores of Rs.

Months	1934-35	1935-36	1936-37
April ..	10.9	11.6	10.1
May ..	10.5	11.8	10.0
June ..	9.6	9.9	9.8
July ..	9.9	10.1	10.1
August ..	10.7	11.2	9.3
September ..	10.5	10.2	9.5
October ..	12.5	11.8	10.7

Months	1934-35	1935-36	1936-37
November	11.4	12.7	10.6
December	10.3	10.6	9.9
January	12.9	13.1	12.6
February	10.6	10.5	9.3
March	12.4	10.8	13.1

Calculate —

(a) The average import into India for each month for the whole period.

(b) The geometric mean for 1934-35

(c) The median for 1936-37

(56) Find out the average height of a clerk in a certain office from the following figures What is the median height and how far does it differ from the mode?

Height	Frequency	Height	Frequency
5′ 6″	1	5′ 11″	1
5′ 7″	2	6′ 0″	2
5′ 8″	4	6′ 1″	1
5′ 9″	3	6′ 2″	1
5′ 10″	2	6′ 3″	1

(M.A. Agra 1937)

(57) A railway train runs for 50 minutes at a speed of 40 miles an hour and then because of repairs of the track runs for 10 minutes at a speed of 10 miles an hour What is its average speed?

(58) The following is the age distribution of candidates appearing at the Matriculat on and Intermediate Arts examinations of the Patna University in 1937

Age in years	12	13-	14	15-	16-	17	18-	19	20-	21-	22-	Total
Matriculat on	5	48	189	303	522	980	981	794	515	474	×	4811
Intermed ate	×	×	×	5	45	87	127	150	155	127	175	871

Compare the median and modal ages of the Matriculation candidates with those of I A candidates

(M.A, Patna 1940)

(59) Amend the following table, and locate the median from the amended table. Also measure the magnitude of the median so located.

Sizes	Frequency
10—15	10
15—17.5	15
17.5—20	17
22—30	25
30—35	28
35—40	30
45 & upwards	40

(B.Com., Alld., 1942).

(60) Monthly incomes of twenty families are given below in rupees:—

2,000; 35; 400; 15; 40; 1,500: 300; 6; 90; 250; 20; 12; 450; 10; 150; 8; 25; 30; 1,200; 60.

Calculate the Geometric Mean and the Harmonic Mean of the above incomes.

(B.Com., Alld., 1941).

CHAPTER XI

DISPERSION AND SKEWNESS

DISPERSION

Meaning of Dispersion.

Averages of the first order, discussed in the last Chapter, consider only the central position of a series. They do not throw light on the formation of the series. They fail to characterize the detail from which they are made up. Hardly ever are the various items of a series equal to the value of the average computed from them. Some measure of the differences of the items from their average is necessary. Averages of the second order' provide this measure. An average of the second order is an average of the differences of all the items of the series from an average of those items. In averaging these differences their irregularities are brushed off and a representative figure results.

All frequency distributions are not similar. They may differ in two ways. Firstly, they may differ in the numerical size of their averages though not necessarily in their formations. For instance the averages of two series—1 2 3 and 7 8 9—are respectively 2 and 8 and therefore different, but the scatteredness of the items in the two series around their rehpective means is similar, the greatest deviation from the average in either series is only 1. The two series are therefore not differently formed. Secondly, frequency distributions may have the same values of their averages yet differ in their respective formations. Let us suppose that the daily earnings of A and B two mechanics during the six days of a working week are as given below:

Days	A's earnings Rs.	B's earnings Rs.
Monday	4	3
Tuesday	4	4
Wednesday	5	5
Thursday	5	5
Friday	6	6
Saturday	6	7
	Rs 30	Rs 30

The total earning made in 6 days, Rs. 30, and therefore, the average earnings, Rs. 5 per day, are thus exactly the same in both the cases, but the scatteredness of the values of the items of the series round their average earnings is different in the two cases. The greatest deviation from the average in A's case is Re. 1 and in B's case Rs. 2. The two series are, therefore, differently constituted, though they result in averages of the same numerical value. It follows that averages must be used with great caution. To these cautions belongs the measurement of the dispersion or scatteredness of the series around the mean. The value of the mean depends necessarily on the distribution of the items around it and on its position in the series. An examination of this distribution furnishes us with a valuable supplement to the information given by the mean itself: it tells us how the items comprised in a group vary in size. This helps us in finding the extent to which the average is 'typical'.

The term "Dispersion" is used in two senses is Statistics. One sense is general, implying that within a given group the items are not uniform in their size. That is, they differ in their magnitude. This difference may be great or small. Accordingly, dispersion may be considerable or slight. If the profits of a given number of businesses in the same trade and with the same capital are found to vary between Rs. 11,000 and Rs. 11,020, they are scattered over a small range. That is, they are fairly consistent, or, in other words, their variability is slight. If, however, the profits vary between Rs. 1,100 and Rs. 11,000, consistency is wanting, the range is wide, or, in other words, variability is considerable.

The other sense in which the term dispersion is used is more precise. In this sense it indicates an absolute or relative measure of the differences of the items of a group from some average computed from those items. It may be noted that the difference between the measurements of the value of a variable

and its mean or any other fixed point is technically termed deviation. And, a measure of the deviations of the size of items from an average is called the **Measure of Dispersion.**

Measures of Dispersion

The two senses in which the term dispersion is used are important The first sense points to the limits within which data fall the second sense calls attention to the amount absolute or relative by which the values of the items differ from an average or type The two senses are different. Dispersion in the first sense is indicated by the method of limits where the complete range of the items may be shown Dispersion in the second sense is expressed by the method of averaging differences from a type.

Method of Limits

The most common way of measuring dispersion under this method is that of computing the Range

The Range

Range of dispersion represents the difference between the values of the extremes i e the largest, and the smallest items, of the data under review If in a certain class the height of the shortest student were 4 ft 10 in and that of the tallest 6 ft the range would evidently be 14 inches Range is thus the simplest method of measuring dispersion but it is too indefinite to be used as a practical measure of dispersion since it depends entirely upon the values of the extreme items For instance if a dwarf whose height was only 3 ft. 6 inches were admitted to the clas in the above example the range would suddenly rise to 30 inches while the average height of the class would not be materially affected. There is yet another reason why the range is not a satisfactory measure of dispersion It is that the range does not take into account the distribution of the items within its limits This distribution may vary widely even though the extremes be of the same value We might get the same value

for the range from a symmetrical and a J-shaped (*i e*, asymmetrical) frequency-curve. Clearly, two such distributions cannot be regarded to exhibit the same dispersion. On the other hand, series with differnt maxima and minima may have practically the same distribution or dispersion.

Besides, the range is an absolute measure of dispersion, and, therefore, it is not possible to compare through it the relative dispersion of two series expressed in different units. Absolute dispersion measured, say, in yards is not comparable with dispersion measured in tons. If the absolute dispersions are reduced to relative bases, comparison would be possible. This is done by dividing the range by the sum of the extreme items. The quotient is called the **co-efficient** or **ratio of dispersion.**

In our example of the height of students taken above, the range is 14 inches, and the co-efficient is $\frac{14}{130} = .108$.

Method of Averaging Deviations.

Under this method the most common measures are (1) the Average or Mean Deviation, also called the First Moment of Dispersion. (2) the Standard Deviation, and (3) the Quartile Deviation. These are absolute measures. They are converted into relative measures called coefficients for purposes of rendering comparison between series measured in different units possible.

Mean Deviation and its Co-efficient.

The first moment of dispersion, also called the mean or average deviation, is the arithmetic average of the deviations of the group measured from an average (Median, Mode or Mean) taking all deviations as positive. In other words, **mean deviation is the sum of the deviations from an average divided by the number of items.** It is necessary to treat all deviations as positive, since the sum of the deviations from the arithmetic average taken with minus and plus signs is zero, and that from the median and the mode is nearly zero.

Let d stand for deviation i.e. difference between an individual item and the average and m for the value of an individual item. Then

$d = m - a =$ deviation from the arithmetic average
$d_m = m - M =$ deviation from the median
$d_z = m - Z =$ deviation from the mode

If n be the number of items in a series

$\frac{\Sigma d}{n} =$ The First Moment of Dispersion from the Arithmetic Average represented by δ (Delta)

$\frac{\Sigma d_m}{n} =$ The First Moment of Dispersion from the Median represented by δ_m

$\frac{\Sigma d_z}{n} =$ The First Moment of Dispersion from the Mode, represented by δ_z

It is proper to calculate the mean deviation from the median because the sum of deviations and consequently the mean deviation is *least* when median is chosen as the origin from which deviations are measured. In practice, however it is easier and not un satisfactory to calculate it from the arithmetic average. In case the observations are recorded in groups between different limits mean deviation from the median is difficult to calculate with precision and therefore arithmetic mean rather than the median may be chosen as the origin. It is not common practice to calculate the mean deviation from the mode.

Mean deviation gives us the absolute measure of dispersion. This is one factor that is required for calculating the relative measure of dispersion, called **Mean Co-efficient of Dispersion**. The other factor required is the mean used in the particular case. Thus

$\frac{\delta}{a} =$ Co-efficient of Dispersion from the Arithmetic Average.

$\frac{\delta_m}{M}$ = Co-efficient of Dispersion from the Median, and

$\frac{\delta_z}{Z}$ = Co-efficient of Dispersion from the Mode.

Calculation of Mean Deviation and its Co-efficient.

Example 1. Required to find mean deviation and its co-efficient when individual quantitative observations are given.

Table 20. *Calculation of Mean Deviation of X's Monthly Earnings for a year.*

Months	Monthly Earnings m	Deviations from median (Rs. 42) (+&—signs ignored) d_m
	Rs.	Rs.
1	39	3
2	40	2
3	40	2
4	41	1
5	41	1
6	42	0
7	42	0
8	43	1
9	43	1
10	44	2
11	44	2
12	45	3
$n=12$		Σd_m = Rs. 18

Median or M = Value of $\left(\frac{n+1}{2}\right)^{th}$ item

= Value of 6.5th item

= Rs. 42.

Mean Deviation from the Median, or, $\delta_m = \frac{\Sigma d_m}{n}$

= Rs. $\frac{18}{12}$

= Rs. 1.5

Co-efficient of Dispersion from the Median $= \frac{\delta_m}{M} = \frac{1.5}{42}$

$= .0357$ approximately

Since the arithmetic mean and the median have the same values in this example mean deviation from the arithmetic mean and its co-efficient of dispersion will also have the same values as those from the median. We can now state that the mean or median earnings of X are Rs 42 and the earnings deviate from the mean or median on an average by Rs 1.5

Example 2 Required to find mean deviation and its co-efficient when a discrete series is given

Table 21 *Calculation of Mean Deviation*

Size of item m	Frequency f	Deviation from Median (5) (+&— signs ignored) d_m	Total Deviation (Frequency × deviation) fd_m
2	2	3	6
3	3	2	6
4	5	1	5
5	8	0	0
6	6	1	6
7	4	2	8
8	2	3	6
9	1	4	4
	$n=31$		$\Sigma d_m = 41$

Median or M = the value of $\left(\frac{n+1}{2}\right)^{th}$ item

= the value of 16th item

$= 5$

$$\text{Mean Deviation from Median or } \delta_m = \frac{\Sigma d_m}{n}$$

$$= \frac{41}{31} = 1.32$$

$$\text{Mean Co-efficient of Dispersion} = \frac{\delta_m}{M}$$

$$= \frac{1.32}{5} = .264$$

Example 3. Required to find the mean deviation and its co-efficient when the data are composed of a continuous series, measuring the deviations from the median as well as from the arithmetic average.

Table 22. *Calculation of Mean Deviation of Marks of 60 Students in Economics.*

Marks-group	Mid value m	Frequency f	Deviation from Median (35 marks) (+&—signs ignored) d_m	Total Deviation from Median fd_m	Deviation from Arithmetic Average (34.7 marks) (+&—signs ignored) d	Total Deviation from Arithmetic Average fd
0—10	5	4	30	120	29.7	118.8
10—20	15	8	20	160	19.7	157.6
20—30	25	11	10	110	9.7	106.7
30—40	35	15	0	0	0.3	4.5
40—50	45	11	10	110	10.3	113.3
50—60	55	7	20	140	20.3	142.1
60—70	65	4	30	120	30.3	121.2
		$n=60$		$\Sigma d_m=760$		$\Sigma d=764.2$

(*i*) Median or $M=35$ marks, by interpolation in (30-40) marks group.

$$\text{Mean Deviation from Median or } \delta_m = \frac{\Sigma d_m}{n} = \frac{760}{60} \text{ marks}$$

$$= 12.67 \text{ marks, approx.}$$

Co-efficient of Dispersion from Median $=\frac{\delta_m}{M}=\frac{12\ 67}{35}=36$ approx.

(ii) Arithmetic Average or $a = 34\ 7$ marks

Mean Deviation from Arithmetic Average or $\delta=\frac{\Sigma d}{n}=\frac{764\ 2}{60}$ marks

$=12\ 74$ marks approx

Co-efficient of Dispersion from Arithmetic Average $=\frac{\delta}{a}=\frac{12\ 74}{34\ 7}$

$=37$ approx.

The above example amply demonstrates that mean deviation when measured from the median is *least* for δ_m is less than δ and it would also be less than δ_2 or mean deviation from any other point

Characteristics and Uses of Mean Deviation and its Co-efficient

Mean deviation and mean co-efficient of dispersion are easy to calculate and comprehend they take all items into consideration and give weight to deviations according to their size. The co-efficient is usefully employed in economic studies like the distribution of personal wealth in a community where the rich and the poor both are considered But the mean deviation does not lend itself readily to algebraical treatment Other moments of dispersion have therefore come into use.

Standard Deviation and its Co-efficient.

An alternative method of eliminating the algebraical signs of the deviations from an average is to square up each deviation This method is employed here **Second moment of dispersion** is the sum of the squares of the individual deviations

from the arithmetic average divided by their number, *viz*, $\frac{\Sigma d^2}{n}$ **Standard deviation is the square root of the Second Moment,** *viz*, $\sqrt{\text{Second Moment}}$ or $\sqrt{\frac{\Sigma d^2}{n}}$ -represented by σ (Sigma). Standard deviation is an absolute measure of dispersion and is invariably computed from the arithmetic average, since it is *least* when arithmetic average is chosen as the origin from which deviations are measured. To compute the **Standard Co-efficient of Dispersion,** a relative measure, the standard deviation is divided by the arithmetic average,

viz, $\frac{\sigma}{a}$

Similar moments and absolute measures of dispersion based on mode and median are quite conceivable. Second moment of dispersion computed from any mean or value other than the arithmetic average is sometimes termed **Mean Square Deviation** and the absolute measure of dispersion based on such second moment is designated **Root-Mean Square Deviation** But root-mean square deviation is converted into standard deviation before any use is made of it. Since the sum of the squares of the deviations from the arithmetic average is minimum, it is obvious that standard deviation is that root-mean square deviation whose value is the least, and second moment is that mean square deviation whose value is minimum.

Calculation of Standard Deviation and its Co-efficient.

Direct Method

Example 4. Required to find standard deviation and its co-efficient when quantitative individual observations are given.

Table 23 Calculation of Standard Deviation of X's Monthly Earnings for 12 months by the Direct Method

Months	Monthly earnings	Deviation from Arithmetic Average (Rs 42)	Square of Deviation
	m	d	d^2
1	39	−3	9
2	40	−2	4
3	40	−2	4
4	41	−1	1
5	41	−1	1
6	42	0	0
7	42	0	0
8	43	+1	1
9	43	+1	1
10	44	+2	4
11	44	+2	4
12	45	+3	9
$n=12$	$\Sigma m=$ Rs 504 $a=$ Rs 42		$\Sigma d^2=38$

Arithmetic Average or $a=\frac{\Sigma m}{n}=$ Rs $\frac{504}{12}=$ Rs 42

Second Moment of Dispersion $=\frac{\Sigma d^2}{n}=\frac{38}{12}=3.17$ Sq Rs approx

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{3.17}=$ Rs 1.78 approx

Standard Co-efficient of Deviation $=\frac{\sigma}{a}=\frac{1.78}{42}=.042$ approx

We can now state that the arithmetic average of the given series is Rs 42 and its standard deviation is Rs 1.78

Example 5 Required to find standard deviation and its co-efficient when a discrete series is given

Table 24. *Calculation of Standard Deviation by the Direct Method.*

Size of items m	Frequency f	Sum of sizes (Frequency × size of item) mf	Deviation from mean (8) d	Square of deviation d^2	Product of square of deviation and frequency fd^2
2	1	2	−6	36	36
4	2	8	−4	16	32
6	3	18	−2	4	12
8	5	40	0	0	0
10	3	30	+2	4	12
12	2	24	+4	16	32
14	1	14	+6	36	36
	$n=17$	$\Sigma m=136$ $a=8$			$\Sigma d^2=160$

Arithmetic Average or $a=\frac{\Sigma m}{n}=\frac{136}{17}=8$

Second Moment of Dispersion $=\frac{\Sigma d^2}{n}=\frac{160}{17}=9.4$ approx.

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma d^2}{n}}\sqrt{9.4}=3.066$ approx.

Standard Co-efficient of Deviation$=\frac{a}{\sigma}=\frac{3.066}{8}=.383$ approx.

We can state that the arithmetic average of the given series is 8, and its standard deviation is 3.066.

Example 6. Required to find the standard deviation and its co-efficient when a continuous series is given.

Suppose the class-intervals of a given series are 1-3, 3-5, 5-7, 7-9, 9-11, 11-13, 13-15 and their respective frequencies are 1, 2, 3, 5, 3, 2, 1. Then, the class-intervals will be placed in the first column of a table, their mid-points 2, 4, 6, 8, 10, 12, 14, in the second column, which tally with the size of items of table 24.

example 5 above The entire working of the example will therefore be the same as that in example 5 resulting in this particular case in the same values of arithmetic average, second moment standard deviation and standard co-efficient of dispersion

Short Cut Method

The preceding direct method of computing the standard deviation is easy and simple if the arithmetic average of the series is a whole number as it was in examples 4 5 and 6 above *But often the arithmetic average happens to be a fraction.* Then the deviations from the true mean would also contain fraction The calculation of such deviations and their squaring up will be tedious In such cases the short-cut method of computing the standard deviation can be usefully employed in place of the direct method

The short-cut method of computing the standard deviation, as we shall presently see saves valuable time and tiresome effort. It gives the same value for the standard deviation as the direct method does The procedure is as follows Assume any size of item as an average compute deviations from it square each deviation summate, subtract from the summation n times the square of the difference between the assumed average and the true average divide by n and extract the square root of the quotient. Symbolically

$$\sigma = \sqrt{\frac{\Sigma d^2_x - n\,(a-x)^2}{n}} \qquad \text{or} \quad \sigma = \sqrt{\frac{\Sigma d^2_x}{n} - (a-x)^2}$$

where x represents assumed average[1] a actual average d^2_x square of deviation from assumed average and n number of observations

Example 7 Required to calculate standard deviation of a continuous series The procedure worked out in this example shall also apply to series of any other type)

[1] The assumed average x is a whole number approximating the actual average That value which has the maximum *frequency* may also be taken as the assumed average or the working mean as it is often called.

Table 25. *Calculation of Standard Deviation by the Short-cut Method.*

(1) Size of item	(2) Mid-value m	(3) Frequency f	(4) Sum of sizes [col. (2) × col. (3)] mf	(5) Deviation from the assumed average (5) d_x	(6) Square of Dev. d^2_x	(7) Product of Square of Dev. & Frequency [col. (3) × col. (6)] fd^2_x
1.5—2.5	2	3	6	—3	9	27
2.5—3.5	3	4	12	—2	4	16
3.5—4.5	4	5	20	—1	1	5
4.5—5.5	5	8	40	0	0	0
5.5—6.5	6	7	42	+1	1	7
6.5—7.5	7	6	42	+2	4	24
7.5—8.5	8	3	24	+3	9	27
		$n=36$	$\Sigma m=186$			$\Sigma d^2=106$

True Arithmetic Average or $a=\frac{\Sigma m}{n}=\frac{186}{36}=5.17$ approximately

Let the Assumed Average or $x=5$; $(a-x)^2=\left(\frac{186}{36}-5\right)^2=\frac{1}{35}$

Standard Deviation or, $\sigma=\sqrt{\frac{\Sigma d^2_x-n\quad(a-x)^2}{n}}$

$$\sqrt{\frac{106-36(\frac{1}{35})}{36}}=1.71 \text{ approx.}$$

Standard Co-efficient of Dispersion$=\frac{\sigma}{a}\ \frac{1.71}{5.17}=.33$ approx.

Characteristics and Uses of Standard Deviation and its Co-efficient.

Standard deviation and its co-efficient possess all those properties which a good measure of dispersion should. The process of squaring the deviations eliminates negative signs, and thus

makes mathematical manipulation of figures easy Largely for this property standard deviation has been used by biologists It has not found its favourites among economists for two main reasons Firstly the squaring of deviations gives greater weight to extreme items than it does to those differing only slightly from the arithmetic average This factor has hardly any value in economic studies since the commercial or economic statistician is interested in the results of the modal class Secondly the computation of standard deviation requires considerably greater time and effort than that of mean deviation For the businessman rapidity in the preparation of results is an important factor Therefore in economic and commercial studies there is a tendency to use mean deviation unless there is a particular reason for using the standard deviation

But standard deviation enjoys at least two decided advantages over the mean deviation Firstly it is in general less affected by fluctuations of sampling and secondly it is more easily amenable to algebraical processes These two properties make standard deviation useful for advanced work Its use is therefore increasing for measuring variability The standard deviation for instance has a special use in the computation of Karl Pearson s Co-efficient of Correlation which will be discussed in a later chapter

Modulus

It is another measure of dispersion based on the second moment of dispersion It is generally represented by c The formula used is $c = \sqrt{\frac{2\Sigma d^2}{n}}$

Variance

The quantity σ^2 is known as Variance

Co-efficient of Variation

According to Karl Pearson who first suggested its use co-efficient of variation is the percentage variation in the mean,

the standard deviation being treated as the total variation in the mean.' In other words,

$$\text{Co-efficient of Variation or } v = 100 \; \frac{\text{Standard Deviation}}{\text{Arithmetic Average}}$$

$$= 100 \times \text{Co-efficient of Standard Deviation.}$$

Thus, in example 8, co-efficient of variation $= 100 \times .33 = 33$. Co-efficient of variation is a relative measure of dispersion and has come into use largely.

Quartile Deviation and its Co-efficient.

The measures of dispersion discussed under the First and the Second Moments of Dispersion take into account the deviations of all the items. **Quartile deviation,** also called the **Semi-Interquartile Range,** is based on the quartiles, and affords a genaral idea of the dispersion of a group without considering each particular item. The first and the third quartiles include between them the middle-half of the items of a group. If the dispersion of this half could fairly represent the dispersion of the whole group, a simple method of measuring it will be:

$$\text{Quartile Deviation or } Q.\,D = \frac{Q_3 - Q_1}{2}$$

where, Q_3 stands for 3rd or upper quartile, and Q_1 stands for 1st or lower quartile.

Quartile deviation is an absolute measure of dispersion. Its relative measure, that is, **Quartile Co-efficient of Dispersion,** is quartile deviation divided by the average of the two quartiles. Symbolically,

$$\text{Quartile Co-efficient of Dispersion} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 - Q_1}{2}} = \frac{Q_3 + Q_1}{Q_3 + Q_1}$$

Calculation of Quartile Deviation and its Co-efficient

Example 9 In example 1 table 20 page 193 Q_3=Rs 43.75 and Q_1=Rs 40.25

$$\text{Quartile Deviation} = \text{Rs } \frac{43.75-40.25}{2} = \text{Rs } 1.75$$

$$\textit{Quartile Co-efficient of Dispersion} = \frac{43.75-40.25}{43.75+40.25}$$

= 0.042 approximately

Upon the assumption that median lies half way between the upper and the lower quartiles we may observe that in our example

$$\text{Median} = \frac{Q_3+Q_1}{2} = \text{Rs } \frac{43.75+40.25}{2} = \text{Rs } 42$$

and the difference occurring on either side of it is Rs 1.75 In other words median is Rs 42 and half the items are within the range Rs 42±1.75

Characteristics and Uses of Quartile Deviation and its *Co-efficient*.

Quartile deviation and its co-efficient are simple to comprehend and easy to compute They are quite satisfactory if one is concerned with the main body—the middle half—of the series and cares nothing about extreme variations Quartile deviation has a serious drawback in that its value will be the same for those series whose quartiles are the same whatever the distribution of the observations *between the quartiles and beyond them* This is because it considers only the quartiles and not each particular item of the array It is therefore not so sensitive as the mean and the standard deviations

Choice of Measures of Dispersion.

A good measure of dispersion should possess some such qualities as an ideal average does *That is it should be based*

on all the observations made, easily calculated, readily understood, affected very little by fluctuations of sampling, and amenable to algebraical treatment. The range, it has already been seen, is not a satisfactory measure, and its co-efficient, therefore, is not very much favoured by statisticians. Quartile deviation enjoys two advantages over the standard and mean deviations: It is easier to calculate and clearer in meaning. But, since it has no simple algebraical properties and is liable to be erratic, it is not good for any but the most elementary statistical work, where only a rough estimate is desired. It is, however, not unsatisfactory when the distribution of values in a series is fairly symmetrical. If the distribution lacks symmetry and there are great differences in frequency between successive values of items in the series, it is better to select measures which give each value its due weight. Such measures are the mean deviation and the standard deviation. Mean deviation is less troublesome to calculate than the standard deviation, but cannot be used for further mathematical operations. If in a given problem median suits the best, mean deviation would be a good measure of dispersion. But as arithmetic mean is the most commonly used average, standard deviation, which is invariably computed from the arithmetic average, is the most important and the best measure of dispersion. More so, because it is the least erratic, and is suitable for further algebraic manipulation. Its use is, therefore, recommended for cases where positive and comparatively precise results are desired.

Absolute and Relative Measures of Dispersion.

Two or more groups may be compared by stating their respective means and absolute measures of dispersion provided the means of the groups do not vary greatly in size and the groups are measured in the same unit. If the difference between the means is great, it is safer to compare their relative measures of dispersion, *i.e.*, co-efficient, rather than absolute measures. For example, if the production of a commodity in factory A for

three successive years be 1,250, 2000 and 2750 units respectively and that of factory B for the same period runs in the order 3250, 4000 and 4,750 units, the mean deviation 500 units is exactly the same in both cases. Similarly are the ranges and the standard deviations identical. If only these absolute measures are compared a fallacious conclusion that the variability of production in the two factories is the same might be drawn. But when the mean co-efficient of dispersion for the production of factory A .25 is compared with the mean co-efficient of dispersion for factory B .125, the difference in the degree of variability in the production of the two factories is made clear. According to the general rule that **the lower the co-efficient of dispersion** (mean, standard or quartile) **the smaller is the variability of the series**, or in other words the greater is its consistency, production in factory A is more variable than that in factory B. Co-efficients of dispersion therefore correct the wrong impression created by the absolute measures.

The absolute measures are concrete quantities and should be stated in terms of the units of the variable (Rs., miles, inches, years, etc.). It is impossible to compare dispersions in two universes measured in different units—say, one measured in Rs. and the other in yards—through absolute measures of dispersion. For this reason also, as already pointed out, co-efficients of dispersion are computed. They enable comparison between universes of different characters since they are pure numbers.

Relation between Measures of Dispersion

Mean, standard and quartile deviations all measure the same property of dispersion but they do it in different ways. There does not exist a perfectly definite relation between them. Yet for moderately symmetrical distributions the following relations are *approximately* true:

(1) Quartile Deviation $= \frac{2}{3}$ (Standard Deviation)

(2) Mean Deviation$= \frac{4}{5}$ (Standard Deviation), when standard deviation is measured from the arithmetic average, a.

(3) $a \pm 2\sigma$ or $a \pm 3\sigma$ shall cover a majority of the observations of the group.

The above relations are not likely to hold good in cases where the number of observations is comparatively small.

Lorenz Curve.

The graphic method can also be used for showing the dispersion of a group. The method adopted by Dr. Lorenz for the study of the distribution of wealth (the curve showing the distribution being designated after him as Lorenz Curve) is the best for the purpose. It is, in effect, a cumulative percentage curve, combining the percentage of items under review with the percentage of the factor (say, wealth) distributed among the

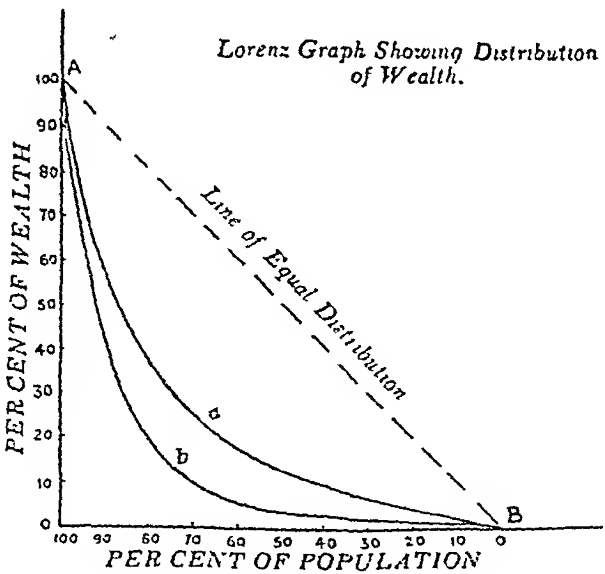


Fig 1.

items In other words it is an ogive curve* formed by cumulating the values of the items on each axis and reducing the values thus obtained to percentages of the total Figure 1 on the last page illustrates the construction of the curve If wealth were equally distributed among the people the curve would be a straight line AB connecting the two extremes of the scales In practice however curves like *a* or *b* result The less the distance between the curve showing actual distribution and the line of equal distribution the greater is the homogeneity in the distribution of wealth or less is the dispersion While the farther the curve of actual distribution from the line of equal distribution the larger is the percentage of poor people and the greater is the concentration of wealth in the hands of a few millionaires It is evident that dispersion in *b* is greater than that in *a* in Figure 1

The Lorenz Curve does not yield a numerical measurement of dispersion and is in this respect inferior to measures of dispersion But its merit is that it affords a picture of dispersion at a glance It should be used along with a co-efficient of dispersion when a detailed study of dispersion is desired Lorenz Curve is very useful in such studies as the distribution of land wages and income among the population of a country or the distribution of profits over different groups of businesses.

Practical Utility of Measures of Dispersion

Measures of dispersion it has already been indicated supplement the information given by the mean They are computed to determine the typical character of an average They may also be computed for estimating the value of a series itself For instance the dispersion of time series affords a measure of the consistency or variability of the phenomenon to which the series relates Determination of the degree of variability is at times very important in politico economic problems. Violent fluctuations in production or trade naturally give a shock to the

* See Chapter XV

economic organism, which affects many people. Measures of dispersion enlighten us on the degree of these variations. Measures of dispersion are invaluable in the study of such problems as inequalities of income in a country, distribution of land among different units of agricultural community, wage fluctuations, etc. They enable comparison between different groups of phenomena. which is an important function that the Science of Statistics has to perform.

SKEWNESS

Skewness denotes the opposite of symmetry. As applied to frequency distribution it indicates that the distribution of items in it is not symmetrical. Skewness relates to the shape of the curve of a frequency distribution.

Tests of Skewness.

The presence of skewness or asymmetry in a given series can be tested in several ways. In a symmetrical series the mode, the median and the arithmetic average are identical. Therefore, skewness or lack of symmetry in a series is shown when these three averages do not coincide. It is further shown when the sum of the positive deviations from the median in a series is not equal to the sum of the negative deviations. It is also found to be existing when the quartiles, or pairs of deciles, are not equidistant from the median. It is also shown when at points of equal deviations on either side of the mode the frequencies are unequal. Lastly, if skewness is present in a frequency distribution, its graph will not give the normal, bell-shaped, symmetrical form. Rather, the base would be stretched to a greater extent on one side than on the other. In curves which are not far away from being symmetrical the median usually covers over two-thirds of the distance travelled by the arithmetic average from the mode (See Figure 2, page 215). Therefore, approximately,

$$M = Z + \frac{2}{3}(a - Z)$$

where M stands for median Z for mode and $\bar{x}$ for mean. In a skew curve mode median and arithmetic average normally occur in sequence the last being pulled away the farthest in the direction in which the curve is skewed.

We shall apply the above tests to the following example.

Table 26. *Calculation of Arithmetic Average Median and Mode and Standard Deviations*

1	2	3	4	5	6	7	8	9
Size of item m	Frequency f	Cumulative frequency cf	Total size of items mf	Dev. from the median and mode d_m d_o	Total Dev. from median and mode fd_m fd_o	Dev. from mean d	Square of deviation d^2	Total squares of deviation fd^2
3	1	1	3	-4	4	-4.1	16.81	16.81
4	2	3	8	-3	6	-3.1	9.61	19.22
5	3	6	15	-2	6	-2.1	4.41	13.23
6	5	11	30	-1	5	-1.1	1.21	6.05
7	8	19	56	0	0	-0.1	.01	.08
8	6	25	48	$+1$	6	0.9	.81	4.86
9	4	29	36	$+2$	8	1.9	3.61	14.44
10	2	31	20	$+3$	6	2.9	8.41	16.82
11	1	32	11	$+4$	4	3.9	15.21	15.21
		$n=32$	$\Sigma m=227$	-10 $+10$	45	10.5 $+9.6$		$\Sigma fd^2=106.7$

First Test. The arithmetic average is $\frac{227}{32}$ or 7.1 the median—value of 16.5th item—is 7 and the mode is also 7. The values of mean median and mode therefore do not exactly coincide. The curve is not perfectly symmetrical.

Second Test. Deviations from the median in column 5 show fine symmetry. The negative sum of deviations -10 is equal to the positive sum $+10$ leading us to think that the curve is symmetrical.

Third Test. $Q_3 - M = 8 - 7 = 1$, $M - Q_1 = 7 - 6 = 1$. The two quartiles happen to be equidistant from the median. Therefore the curve of the series appears to have a symmetrical form.

Fourth Test: We take equal deviations on either side of the mode and compare the frequencies centering round them. 6—and 8 are equidistant from the mode which is 7. But the frequencies against 6 are 5 and against 8 are 6. Therefore, the series is not perfectly symmetrical.

The first and the fourth tests indicate that the shape of the curve of the given frequency distribution lacks symmetry; while, the second and the third tests show that the curve might be symmetrical. The second and the third tests do not always provide a correct answer. We may, therefore, conclude that the curve of the given series falls slightly short of perfect symmetry. '*By how much does the curve lack symmetry*'? may be the natural question. For answering it, we must reduce skewness to numerical quantity. This brings us to the measures of skewness.

Measures of Skewness.

The first measure of skewness, and the simplest too, is based on the fact that in a skew distribution the mean and the mode are drawn widely apart. The larger the distance that the mean (a) is pulled beyond the mode (Z), the greater is the degree of skewness. The **second measure of skewness** is based on the fact that in a skew curve the median does not lie half-way between the quartiles, the quartile nearest to the stretched base being pulled in that direction more than the other quartile. But these measures of skewness should be reduced to **co-efficients of skewness** for just the same reasons as measures of dispersion are reduced to co-efficients. In computing the co-efficients of dispersion the measures of absolute dispersion are divided by the average used. Average would not be the proper divisor here, for the question now is not how much the curve is asymmetrical in proportion to values of the items of the series, but how much more the items on one side deviate than they do on the other. Therefore, measures of skewness shall be divided by the related

measures of dispersion On these principles are based the measures and co efficients of skewness discussed below

First Measure and Co-efficient of Skewness

Measure of Skewness $= a-Z$ i e the difference between mean and mode

Co-efficient of Skewness or $j = \dfrac{a-Z}{\delta_x}$ (A)

or $j = \dfrac{a-Z}{\delta}$ (B)

And if the mode is ill defined the numerator may be substituted by the difference between mean and median so that

$$j = \frac{a-M}{\delta_m} \qquad \text{(C)}$$

Karl Pearson has given a formula in which standard deviation is employed as the denominator rather than mean deviation used above so that

$$j = \frac{a-Z}{\sigma} \qquad \text{(D)}$$

And if the mode is ill defined then on the basis of the relationship between median mode and mean pointed out above[a] formula (D) may be modified as below

$$j = \frac{3\,(a-M)}{\sigma} \qquad \text{(E)}$$

The above co-efficients yield a pure number and are therefore, independent of the units in which the variable is measured and secondly they shall give a zero for symmetrical series These are the two properties which a good measure of skewness should possess It is why the above methods of measuring skewness are termed ideal or standard methods

According to these formulæ the following are the results of our series The measure of skewness $= a-Z = 7{\cdot}1-7 = +{\cdot}1$

$$j = \frac{a-Z}{\delta_x} = \frac{{\cdot}1}{1{\cdot}406} = +{\cdot}071 \quad j = \frac{a-Z}{\delta} = \frac{{\cdot}1}{1{\cdot}425} = +{\cdot}0702$$

[a] See page 209

$$= \frac{a-M}{\delta_m} = \frac{7.1-7}{1.406} = +.071 \quad j = \frac{a-Z}{\sigma} = \frac{.1}{1.825} = +.0548$$

$$j = \frac{3(a-M)}{\sigma} = +.1644$$

In theory there is no limit to the values of co-efficients yielded by the formulæ A, B, C, & D, and this is a slight drawback. In practice the results are rarely very high, and for moderately asymmetrical curves they are usually less than unity. The co-efficient yielded by the formula E lies between the limits—3 and+3. In practice it rarely approaches that limit.

Since none of the above formulæ yields zero co-efficient of skewness for our series, we are led to conclude that the curve is skew. But, since the value of the co-efficient is very small, we may add that the degree of skewness is very slight. Skewness, it should be noted, is positive in this case, *i e.*, the mean is greater than the mode.

Second Measure and Co-efficient of Skewness.

The Measure of Skewness$=(Q_3-M)-(M-Q_1)$

$$=Q_3+Q_1-2M$$

where, Q_3 stands for upper quartile, Q_1 for lower quartile and M for median.

The Co-efficient of Skewness, or $j = \dfrac{(Q_3-M)-(M-Q_1)}{(Q_3-M)+(M-Q_1)}$

$$= \frac{Q_2+Q_1-2M}{Q_3-Q_1} \quad \ldots\ldots(F)$$

This co-efficient is also a pure number, and is zero for symmetrical distributions. Its result varies from —1 to +1. The fact whether the particular value of a co-efficient is significant or not is a matter of experience. It may, however, be suggested that 1 denotes a moderate degree and .3 a considerable degree of skewness.

According to the above formulæ, in our series, table 26, page 210

$$\text{Measure of Skewness} = Q_3 + Q_1 - 2M$$
$$= 8 + 6 - 14$$
$$= 0$$

$$\text{Co-efficient of Skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$
$$= \frac{0}{2}$$
$$= 0$$

This measure of skewness and co-efficient suggest that the curve of the frequency distribution in table 26 is symmetrical—a suggestion not warranted by the first measure and co-efficient of skewness. This is the weakness of the second measure and co-efficient. It is the same weakness as that possessed by quartile co-efficient of dispersion—that is, it fails to take into account the size of the extreme variations since it is concerned only with the quartiles and the median. This limitation should be clearly borne in mind before results yielded by this formula are relied upon. This co-efficient is simple and easy to calculate and is sufficiently reliable in those studies in which extreme instances are considered unimportant. It is a rather rough and ready measure and might be used where quartile deviation is being used as a measure of dispersion. Where comparatively greater accuracy is required, Karl Pearson's Co-efficient of Skewness should be employed.

Positive and Negative Skewness.

Skewness can be positive as well as negative. If the arithmetic average is greater than the mode or the median, skewness is positive. If it is less, skewness is negative. When skewness is positive mean would travel to the right of the mode in the curve would be to the left of the mode when skewness is negative our example, the answer is positive. Positive and negative wness are symbolised by plus and minus signs respectively.

Figure 2, below, illustrates slight *positive* skewness. It also shows the positions of the mode, the median and the arithmetic mean, indicated by *Z*, *M*, and *a*-respectively, in an ideal moderately asymmetrical distribution: the median travels about 2/3rds the distance travelled by the mean from the mode. It is also clear that the mode, the median and the mean occur in sequence in a skew curve, and the mean is pulled away the largest in the direction in which the curve is skewed. It may also be noted that the median bisects the area of the curve (called histogram). Mode will remain

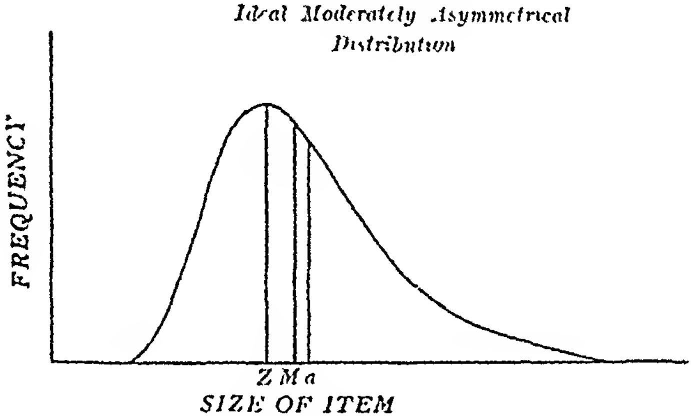


Fig 2

unaffected by the addition of a few more items, but the median and the mean will be deflected.

Dispersion and Skewness Contrasted.

Measures and co-efficients of dispersion, respectively, indicate the absolute and relative differences between the individual items of the series and an average taken as the standard. They do not, however, show the extent to which deviations cluster above or below the average selected.

Measures of skewness, on the other hand, show the extent to which distributions are pulled away, or distorted, from the ideal, symmetrical curve. In a symmetrical curve the mode,

the median and the mean coincide in an unsymmetrical curve they do not Measures of skewness have two functions to perform firstly, they indicate the direction of asymmetry through their positive and negative character secondly they measure the amount of asymmetry in absolute or relative terms through the va i obtained for the measure or the coefficient

The theory of skewness is more important in biological studies and other stud es depending more or less upon the laboratory experiments than in economic and social investigations *In social and economic inquiries a perfectly symmetrical distribution is an exception* and a large degree of skewness is generally expected

It is interesting to note the important part that median and quartiles play in statistics Three characteristics of a group can be studied through them alone the median locating th central position quartile deviation showing dispersion and the second measure of skewness showing skewness It m y again be noted that skewness relates to the shape of the curve rather than to its size

EXERCISES

(1) What do you understand by dispersion? Explain the various methods of its measurement and point out their advantages and disadvantages

(B Com Luck 1030)

(2) Mention the important methods of measuring disper sion and discuss their comparative merits

(B Com Agra, 1943)

(3) What is Skewness? How would you find it in a non symmetrical distribution? Distinguish between positive and negative skewness

(4) What is meant by Skewness? How does it differ from dispersion? What is the object of measuring these?

(B Com, Alld, 1943)

(5) Distinguish between absolute and relative measures of dispersion. Why are the latter computed?

(6) Write short notes on—

Range, First Moment and Second Moment of Dispersion, Standard deviation and Quartile deviation.

(7) Describe the methods of calculating the Standard Deviation and state the relationship between it and the Mean Deviation for a moderately asymmetrical distribution.

(8) What do you understand by Modulus, Variance and Co-efficient of Variation? Give their formulæ.

(9) Explain with the help of specimen curves

(*a*) Lorenz Curve, (*b*) Moderately asymmetrical curve. Point out their salient features.

(10) Calculate the Mean and the Standard Co-efficients of Dispersion of the two series relating to marks in Economics and Politics, given in exercise 23, Chapter X, page 174. What light do the coefficients throw on the variability of the series?

(11) Find the Mean and Quartile deviations of the population of 36 cities of India given in exercise 20, Chapter X, page 174.

(12) Find the coefficient of skewness of the series given in exercise 22, Chapter X, page 174. What is the character of skewness—positive or negative? What does it imply?

(13) From the data given in exercise 24, Chapter X, page 175, compute the coefficient of skewness, and state what light the coefficient throws on he shape of the curve to be drawn from the data.

(14) Calculate the mean and standard deviations of the marks obtained by students in Class A and Class B, given in exercise 29, Chapter X, page 177. State, what you can, about the variability of the marks.

(15) Find the standard deviation and its coefficient for the frequency distribution given in exercise 36, Chapter X, page 179.

(16) Find the mean deviation and its coefficient for the data given in exercise 36, Chapter X.

(17) Draw a curve of the figures given in exercise 40, Chapter X, page 180. Do you think it is a perfectly symmetrical curve? Verify your answer by finding the mode, median and mean.

(18) Calculate the coefficient of mean deviation and the coefficient of quartile deviation for the marks obtained by 30 students, given in exercise 46, Chapter X, pages 112-183.

(19) Calculate the range and its coefficient for the values of exports and imports given in exercise 43 Chapter X page 181 Comment on your result.

(20) Calculate the coefficient of variation of the following monthly incomes of twenty families given below in rupees —

2000 35 400 15 40 1500 300 6 90 250 70
12 450 10 150 8 25 35 1200 60

(B Com Alld 1941)

(21) Find the Arithmetic Average the First Moment of Dispersion and the Standard Deviation from the data in the following series —

Size of item	Frequency
3—4	3
4—5	7
5—6	22
6—7	60
7—8	85
8—9	32
9—10	8

(B Com Alld 1942)

(22) The following table shows the number of workers in two factories whose weekly earnings are given in column (1) Determine the mean values of weekly earnings and standard deviation in both factories

Range of weekly earnings	Number of workers in Factory A	Number of workers in Factory B
4—6	4	71
6—8	376	37[illegible]
8—10	344	303
10—12	115	112
12—14	13	18
14—16	6	1
16—18	9	3
18—20	9	[illegible]
	0	1
TOTAL	900	900

(M A Cal 1936)

(23) Calculate the mean deviation from the following data. What light does it throw on the social conditions of the community?

Difference in age between husband and wife in a particular community

Difference in years	Frequency
0—5	449
5—10	705
10—15	507
15—20	281
20—25	104
25—30	52
30—35	10
35—40	4

(B.Com., Bom., 1956).

(24) What do you understand by the term 'dispersion'? Name and describe two standard measures of dispersion.

This index numbers of prices of cotton and coal shares in 1942 were as under :—

Month	Index number of prices of cotton shares	Index number of prices of coal shares
January	188	131
February	178	130
March	173	130
April	164	129
May	172	129
June	183	120
July	184	127
August	185	127
September	211	130
October	217	137
November	232	140
December	240	142

Which of the two shares do you consider more variable in price?

(M A , Agra, 1944)

(25) Calculate the mean and the standard deviations of the following figures and state the percentage of cases which lies out-

side the mean at distances $a\pm\sigma$ $a\pm 3a\pm 3\sigma$, where σ stands for the standard deviation

115 117 121 125 116 120 118 117 119 116
122 124 123 118 120 118 126 127 122, 125

(26) The following table gives the exports of some commodities from India —

	1937 38	1938 39	1939-40	1940-41	1941 42
Exports of Pig Iron (000 tons)	6 9	5-5	572	596	522
Exports of Raw Cotton 000 bales)	2730	2703	2948	2168	1438
Exports of Cotton goods (million yds)	241	177	221	390	779

Which of the above exports is the most variable from year to year

(27) *Summary of Receipts and Passengers of a certain Motor Bus Co*

Year	Receipts	Passengers
1925	2 354	50 010
1926	2 780	61 060
1927	3,011	70 005
1928	3 020	70 110
1929	3 541	83 001
1930	4 150	91 100
1931*	5 000	100 000

(B Com Alld 1932)

From the foregoing data find out one measure of dispersion and state whether the variation in receipts is greater than that in passengers

(B Com Alld 1932)

(28) Calculate the Standard Deviation of the following data with regard to 2 298 families in the U K.

* The figures for 1931 are mere estimates

Number of persons in the family	1	2	3	4	5	6	7	8	9	10	11	12	Total
Number of families	166	552	580	433	268	148	77	41	20	8	5	1	2298

(M.A., Alld., 1942).

(29) The following are the rents of 18 houses in a certain locality:—

Rs.	a.	Rs.	a.
6	8	6	4
5	0	3	0
5	4	9	0
5	8	4	8
5	4	4	0
4	12	3	0
4	0	3	12
5	0	5	0
4	8	3	0

Calculate the mean deviation of this group.

(B.Com., Luck., 1930).

(30) The following table gives the number of finished articles turned out per day by different numbers of workers in a factory. Find the mean value and 'standard deviation' of the daily output of finished articles, and explain the significance of 'standard deviation'—

Number of articles	Number of workers	Number of articles	Number of workers
18	3	23	17
19	7	24	13
20	11	25	8
21	14	26	5
22	18	27	4

(B. Com., Cal., 1937).

(31) Write short notes on
(1) Dispersion
(2) Standard deviation.

Calculate the standard deviation from the following data

Size of item	Frequency
6	3
7	6
8	9
9	13
10	8
11	5
12	4
TOTAL	48

(B Com Bom 1936)
(B Com Nagpur 1944)

(32) Find out a coefficient of dispersion and a coefficient of skewness from the following figures and explain them —

Income in Rupees		Number of Persons
1	and not exceeding 9	50
10	19	70
19	28	203
28	37	406
37	46	304
46	55	42
55	64	5

(B Com Nag 1943)

(33) The following table gives the area under cultivation production net export and price of wheat in the U S.A —

Year	Area under cultivation (millions of Acres)	Production (millions of Bushel)	Net export (millions of Bushel)	Price per Bushel in Cents
1890	36 7	449	109	84
1895	38 9	542	130	[illegible]1
1900	49 2	599	2[illegible]1	6[illegible]
1905	46 3	706	101	75
1910	45 8	6[illegible]5	70	91
1915	60 3	1008	[illegible]40	96
1920	62 3	843	313	183
1925	52 4	669	93	144
1930	62 6	886	112	67
1935	51 3	6[illegible]6		83

Compare the coefficient of variation of area under cultivation and net export of wheat for the whole period

(M A Patna 1941)

CHAPTER XII

INDEX NUMBERS

An Index Number is a number which indicates the level of a certain phenomenon at any given date in comparison with the level of the same phenomenon at some standard date. It offers a device for estimating the relative changes of a variable in cases where measurement of its actual changes is inconvenient or impossible. If we want to measure the changes from one period to another in a factor, the change may not be capable of direct measurement. But, an evidence of it may be had from a measurement of the quantities *influenced* by the factor under consideration. These measurements may be expressed in different units. Therefore, their movements will not be directly comparable. To make them comparable, we may reduce the changes to a common denominator. We may, therefore, express them as percentages of similar measurements for a selected date. When so expressed, the percentages shall form a group. Each one of these percentages shall throw some light on the incommensurable, hidden factor about which we desire information. If we take an average of all these percentages, it would afford an approximate idea of the change in the factor in question. This average is called Index Number. Thus, averages linked with percentages constitute the whole basis upon which is raised the superstructure of a simple device of comparing factors which are not *directly* comparable.

Let us take an example. Suppose we are concerned with measuring the general changes in the price level. These changes are not directly measurable. Evidence of them can be seen in changes in the prices of different commodities

Quotations of these prices shall be available in different units e g of wheat in Rs per maund of cotton in Rs per bale of petrol in Rs per gallon They are not directly comparable If we plot them on a graph paper reliable conclusions with regard to their movements will hardly be drawn To make them comparable we may express them as percentages of corresponding prices of some selected date These percentages relating to different commodities for the same date shall constitute a group Each of these percentages shall reflect in one way or the other, the change that has taken place in the general price level When we compute an average of these percentages the resulting average would show an approximate general change in the level of prices from the standard date to the date under consideration This average is called the Wholesale Price Index Number, or the General Index Number

Index numbers are not used for measuring changes in the general level of prices alone They are as well employed to measure movements of wages employment cost of living sales production investment business activity shares and stocks and a multitude of other phenomena over a period of time In fact where an attempt is made to bring to light what is enshrouded in complex variations of the items of a time series they are invaluable to use Movement in prices is a matter of general economic interest To a layman a rupee is just a rupee, but changes in its purchasing power are very often a nuisance The technique of index numbers is partly for this reason generally studied in connection with prices

Fluctuations In General Price Level

Prices of commodities fluctuate very often When the price of a single commodity changes the reason for it may be found in a change in the supply of that commodity without a corresponding change in its demand or in change in its demand without a corresponding change in its supply And,

if the price of a commodity which is a substitute for another changes, the reason for it may, in addition, be attributed to a change in the supply or demand of the substitute. But, when the prices of two, or a number of, unrelated commodities, may brassware and cloth, rise or fall together, several reasons may be advanced to explain the situation, but the real one may lie in an alteration in the measuring rod itself. This measuring rod is money, and its value, unlike that of other measuring instruments such as the foot-scale, the maund and the yard-stick, changes with supply of and demand for it. Change in the value of money implies change in prices. According to the Quantity Theory of Money, a fall in the value of money is the same thing as a rise in the general level of prices, or *vice versa*. But, if change in the value of money were the only cause of change in prices, the prices of all commodities would show nearly the same proportional rise or fall, and, therefore, it would be easy to say in what direction the value of money or its purchasing power moved by looking at the change in the price of any single commodity. As we know, change in the value of money is only one of the several causes, and it may be so intermingled with other causes like fluctuations of demand for and supply of goods that it may be difficult to say to what extent the value of money has changed. This complex phenomenon is simplified through the device of Price Index Numbers.

CONSTRUCTION OF INDEX NUMBERS OF PRICES.

We have seen the necessity of constructing Index Numbers of Prices. The technique of their construction, or, for the matter of that, of the construction of any index number, involves the following major problems:—

I. The selection of items to be included; their number and quotations.

II. The choice of the base period.

III The type of average to be used
IV The system of weighting to be adopted

Selection of Items.

Selecting the Commodities

The general index number is based on the prices of commodities exchanged in the market But, it is neither possible to include nor to obtain regular price quotations of *all* the commodities that are bought and sold in various markets of a country Therefore the number of commodities on whose prices the general index number is based has to be brought down to a manageable limit. That is sample data have to be used The commodities selected for the purpose should be

(a) Representative of the tastes habits and customs of the people,

(b) Easily recognisable and

(c) Unlikely to vary in quality

These restrictions would not allow a large number of manufactured goods to enter the index number, since they vary in quality Nor will personal services be included since they are not represented by tangible goods and can be measured in none but the monetary standard. Reliable quotations however, of foodstuffs raw materials and semi manufactured goods are usually available. Therefore the choice of items on which price index numbers are generally based is restricted to these commodities And even of these commodities those whose qualities and discriptions are standardized are commonly selected

Deciding the Number of Commodities.

The next question that arises is what number of items should be included? There is no hard and fast rule for deciding the number In fact the larger the number of items the better

would be the random sample and the greater would be the tendency for errors to compensate one another. But it should also be noted that complications, expense and delay in constructing the general index number increase with increase in the number of items. Therefore, a reasonable number of items consistent with economy, simplicity and accuracy of construction should be taken. In India, the *Calcutta Wholesale Price Index Number* includes 72 items, and the *Bombay Wholesale Price Index Number* 40. In Britain, the *Board of Trade Wholesale Price Index* includes 200 items, the *Economist* and the *Statist Wholesale Price Indices* include, respectively, 58 and 45 items. In America, a wider range of quotations is in use: the *U.S. Bureau of Labour Statistics' Index of Wholesale Prices* includes 450 items and *Dun's Index* 200.

With these indices may be contrasted the so-called "sensitive" index numbers which are based upon a smaller number of items (say 20) supposed to be specially sensitive to fluctuations in business conditions. 'Index numbers of weekly wholesale prices of certain articles in India', based on 23 commodities compiled by the *Economic Adviser to the Government of India,* the index of 15 commodities compiled by the *Bank of England* and the index of 20 basic commodities compiled by *Federal Reserve Board of New York* may be cited as examples.

Obtaining Price Quotations.

When the commodities have been selected and their number decided, the next task is to **make arrangements for obtaining regular quotations of prices.** Quotations may be had from standard trade journals or from leading businessmen dealing in the commodities selected for the index. Great caution is required in selecting reliable trade journals and dependable business houses. We have seen in Chapter VII[1] why quotations of prices published in India are not fully reliable. If the agency of correspondent businessmen is to be employed, leading

[1] *Cf.* Page 78.

businessmen of one town alone will not be representative of the entire country Nor is it feasible to have quotations for a com modity from leading dealers of all the towns of the country Therefore, representative places from which quotations would be obtained will have to be selected That is, **a sample of towns will be taken.** The criterion would be to select those places where a given commodity is bought and sold in large quantities It is not necessary that quotations for all the selected commodities should be obtained from the same place but it would be econo mical to get quotations for as many representative commodities from one town as possible.

After obtaining a sample of towns it would be necessary to have **a sample of the leading dealers** of the selected towns for the reason that quotations of prices cannot be had from *all* the dealers in selected towns.

How should prices be quoted?—Prices should be quoted as so much money per unit of a commodity (*e g*, Rs 5-8-0 per maund) and not as so much units of commodity per unit of money (*e g* 7 seers 4 chattaks per rupee) The former are called money prices and the latter quantity prices Before 1907, prices in *Prices and Wages in India* published by the Government were expressed as seers and fractions of a seer per rupee. Use of such quantity prices needs proper care and caution Money prices vary inversely with quantity prices and the percentage rise and fall also varies in the two notations. Thus if rice sells at 4 seers per rupee and later changes to 2 seers per rupee the quantity price would be said to show a *fall* from 100% to 50% But the money price for rice would change from Rs 10 per maund to Rs 20 per maund. It would show a *rise* from 100% to 200%

It need hardly be emphasized that **the quotations should relate to wholesale prices** and not to retail prices if changes in the general level of prices are to be measured Wholesale prices are far more uniform over a given region for the sam-

day and are more sensitive to the slightest changes in the conditions of demand and supply than are the retail prices. Wholesale prices are, therefore, a better guide for disclosing movements of economic forces that affect and determine prices than are the retail ones. Retail prices, as a matter of fact, lag behind wholesale prices both in their rise and fall and they also fluctuate between narrower limits.

While making arrangements for obtaining price quotations the **quality of the commodity should be correctly specified,** otherwise prices of different qualities of the same commodities may be quoted from different places at the same time, or from the same places at different times. If such is the case the resulting index would be a hotch-potch, incomparable figure. To make matters easy, quotations of those commodities whose qualities and descriptions are standardised are taken.

If it is found necessary to **give special importance** to a commodity, quotations for a few different qualities of the same commodity may be obtained. For example, if special importance is to be given to sugar, the prices of sugar bearing the trade descriptions of *Marhowrah* Crystal, *Dobarrah,* Java White may be taken separately for each selected town. This would be one way of assigning weights to different commodities in proportion to their importance.

How often should the quotations be obtained?—Quotations can be obtained daily, weekly or fortnightly depending upon the nature of the index number. For a monthly index number two quotations per week would be quite adequate.

A somewhat delicate problem arises when the price of an article is 'controlled' by the government, but illicit sales take place at uncontrolled prices.

Averaging the Price Quotations.

When price quotations have been obtained, they should be averaged. The process would be to add up the prices for

a commodity quoted from all the selected places on a particular date and divide the summation by the number of places The quotient would give the average price for that commodity for the country for the particular date if daily prices are being used or for the week if weekly prices are being received To calculate the average price for the month or the year the procedure would be to summate weekly average or daily average prices of the same commodity and divide the sum by the number of weeks or days as the case may be. If however the index is to be based on the prices of only one town only its wholesale prices would be used and the necessity of striking an average of the prices for the country would be avoided.

The following table gives the average yearly wholesale prices of certain commodities in Kanpur in rupees per maund It is

Table 27 *Average Wholesale Prices of Certain Commodities in Kanpur* 1928—1934

Line	Commodities	Average Prices in Rs per Maund						
		1928	1929	1930	1931	1932	1933	1934
1	Rice	7.3	7.7	5.8	4.1	4.3	4.1	3.7
2	Wheat	7.7	5.5	3.6	2.7	3.4	3.2	2.8
3	Linseed	7.0	8.0	6.5	4.[illegible]	3.5	3.4	3.6
4	Gur	6.5	7.3	6.2	4.2	3.5	3.1	4.1
5	Cotton	34.1	29.8	17.3	13.3	14.8	12.9	13.2
6	Tobacco	17.3	17.1	14.5	11.6	4.9	4.9	5.7

proposed to demonstrate the processes and methods involved in constructing general index numbers from the data given in the table.

Choice of Base.

After the prices have been averaged, the next step in the construction of index numbers would be to reduce the average prices to relatives. For doing so, an appropriate base in terms of which the prices shall be expressed as percentages should be selected. Two methods are available for the purpose:

I. Fixed Base.
II. Chain Base.

Fixed Base Method.

In the fixed base method either (i) the average price of some arbitrarily chosen *year* is taken as the base, or (ii) the average of the prices of a *period* of years is taken as the base, and the base chosen is adhered to for an indefinite time. In following the latter method either the prices for five or ten years may be averaged, or the prices of the entire period for which index numbers are to be constructed may be averaged. This method is useful when the data are reviewed at the expiry of a period of years; but the former should be preferred if the data are to be made of a continuous character. If an arbitrary year is chosen as the base, it may happen to be an abnormal year, for instance, a year of labour unrest, of war or of financial crisis. Therefore, in selecting a base year the fact whether statistics of that year are reasonably normal should be specially considered. If an unusual year is taken as the base, the index numbers calculated on it will have to be qualified with a statement drawing attention to the abnormality of the base year. To avoid all this, a base period is often chosen. In averaging the prices of a group of years, chances of abnormalities being present are reduced. Average of a period of years—rather, of the whole group of years to which the series of prices relate—is representative, is less affected by chance variations and is most generally applicable in statistics. In India, however, the wholesale price indices and most cost of living indices use a single year as base year. Only in some cost of living indices is the average of a few years taken as the base.

Computing Price Relatives.—The average price of th base chosen is taken as 100, and the price in each of the other years is expressed as a percentage of this amount. Thus,

$$\frac{\text{Price of a commodity for the current year}}{\text{Price of the commodity for the base year}} \times 100$$

will give the percentage (or price relative) for the current year Let us choose 1928 as the base year to convert the data relating to other years, in table 27 above, into index numbers The percentage price of rice in 1930 on the basis of 1928 is $\frac{\text{Rs } 5.8}{\text{Rs } 7.3} \times 100 = 79$ This price relative is the index number for rice for 1930 with 1928 as the base. All the relatives in table 28 from line 1 to 6, have been computed in this manner with 1928 as the fixed base year

Table 28 *Fixed Base Relative Index Numbers of Wholesale Prices of Certain Commodities in Kanpur,* 1928—1934 (1928=100)

Line	Commodities	Percentages or Relatives, 1928=100						
		1928	1929	1930	1931	1932	1933	1934
1	Rice	100	105	79	56	59	56	51
2	Wheat	100	71	47	38	44	42	36
3	Linseed	100	114	93	60	50	49	51
4	Gur	100	112	95	65	54	48	63
5	Cotton	100	87	51	39	43	38	39
6	Tobacco	100	100	84	67	29	28	33
7	Total of Relatives	600	589	449	325	278	261	273
8	Average of Relatives	100	98	[illegible]	54	46	44	45
9	Median of Relatives	100	102	82	58	47	45	45
10	Geometric mean of Relatives	100	97	72	53	45	42	44

Chain Base Method.

In the fixed base method the base is fixed in the sense that the relatives for all the years are based on the prices of a single year (1928 in our example) or on the average price of a

period of years. Contrasted with it is the chain base or the shifting base method in which the relatives for each year are calculated upon the prices of the preceding year. Thus, the base year is not fixed, but changes from year to year. If, however, it is desired to convert these relatives to one common base, they are chained together.

Computing Link Relatives.—According to the chain base method, in our example, we would express the 1929 figures as percentages of those for 1928, and get index numbers for the commodities for 1929 on 1928 as base; then, for 1930 we would express the 1930 figures as percentages of those for 1929 and obtain index numbers for 1930 on 1929 as base; and so on. Thus the percentage (or, link relative, as it is called in the case of chain base method) for rice for 1929 is $\frac{\text{Rs. } 7.7}{\text{Rs. } 7.3} \times 100 = 105$, for 1930 is $\frac{\text{Rs. } 5.8}{\text{Rs. } 7.7} \times 100 = 75$, and so on. The link relatives in table 29 from line 1 to 6 are based on the preceding year, that is, the years are linked together.

Table 29. *Chain-Relative, Index Numbers of Wholesale Prices of Certain Commodities in Kanpur,* 1928—1934. (1928=100).

Line	Commodities	Percentages or Relatives Based on Preceding Year						
		1928	1929	1930	1931	1932	1933	1934
1	Rice	100	105	75	71	105	95	90
2	Wheat	100	71	65	75	126	94	88
3	Linseed	100	114	81	65	83	97	106
4	Gur	100	112	85	68	83	88	132
5	Cotton	100	87	58	77	111	87	102
6	Tobacco	100	100	85	80	42	100	116
7	Total of Link Relatives	600	589	449	436	550	561	634
8	Average Link Relatives	100	98	75	73	92	94	106
9	Chain Relatives (1928=100)	100	98	74	54	49	46	49

Type of Average to be used

The relatives arrived at by the fixed base method or the chain base method should be averaged to yield the required final index number In theory, any form of average can be used for the purpose In practice, however we are to choose among (a) arithmetic average (b) median and (c) geometric mean Lines 8 9 and 10 of table 28 give respectively, the arithmetic average median and geometric mean of price relatives computed according to fixed base method These averages are the final index numbers of wholesale prices in Kanpur for different years with 1928 as the base

Arithmetic Mean of Relatives in Fixed Base Method

The arithmetic average has the advantage of being readily intelligible but suffers from a bias which it is not easy to remove It is too much affected by the extremes, it gives too much weight to increasing prices and too little to decreasing ones The arithmetic average of relatives, as we shall see a little later, is not reversible For all these reasons the arithmetic average does not reflect the typical movement of prices

Median of Relatives in Fixed Base Method

Median is the easiest to calculate and enjoys an advantage over the arithmetic average in that it is but little affected by extreme items Median is, therefore, very likely to be more typical of price movements than the mean But it may not be possible to find an *actual* median *e g*, in table 28 the last six medians had to be interpolated. Besides median may be erratic when the number of items is small Again, the median of relatives is not reversible Therefore, median too is not a suitable form of average

Geometric Mean of Relatives in Fixed Base Method.

The Geometric mean is of value when items in a group are considered from the viewpoint of their relative differences rather than that of absolute differences. Therefore, it is reasonable to use it in computing index numbers where the items to the averaged are themselves relatives. It is indeed suitable for measuring the average ratio of change in prices for it gives equal importance to equal ratios of change. For instance, when geometric mean of relatives is taken, the effect of doubling of one price is perfectly counterbalanced by the halving of another. This is not the case with arithmetic average or median. Similarly, if the price of one commodity rises by 50% and that of another falls by 50%, the arithmetic average of relatives will neither rise nor fall implying that there is no change in the price level, while, in fact, both the prices show a change. The geometric mean of the relatives would, in this case, show that there is a change in price. Table 30 illustrates these two ideas.

Table 30 *Fixed Base Index Numbers of X and Y Commodities;* (1941=100).

Commodities	1941 (base year)		1942		1943	
	Price	Relative	Price	Relative	Price	Relative
	Rs.		Rs.		Rs.	
X	5	100	10	200	7.5	150
Y	4	100	2	50	2	50
Total of Relatives		200		250		200
Geometric Mean of Relatives		100		100		87
Arithmetic Average of Relatives		100		125		100

The price of X commodity is double that of 1941 in 1942 and of Y commodity is half that of 1941 in 1942. The arithmetic average index number for 1942 on 1941 is 125 implying a 25% rise in general price level, but the geometric mean index number corrects this impression by showing that there is no change in the level of prices in 1942 as compared with 1941 Again in 1943 price of X commodity has risen 50% over that in 1941 and of Y has fallen 50% below that in 1941 The arithmetic average index number for 1943 on the base 1941 is 100 implying that there is no change in the level of prices but the geometric mean corrects this impression by showing that the index number in 1943 falls to 86 7 or 87 From these examples it will be clear that the geometric mean through its property of giving more weight to small items and less to large ones, creates the effect of reducing the influence of upward movements in prices and increasing that of downward movements This property is of great value in tracing the course of prices. Geometric mean has the further advantage that it makes possible the replacement of commodities which have ceased to be representative by those which have become representative without affecting the balance of the index. Yet another advantage of this mean is that index number calculated by using it is reversible that is change of base year can be made without affecting the proportionate change in the general index. Geometric mean is therefore likely to be more typical of the changes in prices than are the arithmetic mean and the median Its use in index number construction is growing although arithmetic average has so far been largely used

Chain Relatives

In table 29 link relatives from line 1 to 6 have been calculated on the chain base method and totalled up in line 7 In line 8 average link relatives have been computed by dividing the totals

in line 7 by 6, the number of commodities. These are the index numbers on the chain base principle, *i.e.*, the index for each successive year is based on the preceding year. If it is desired to base them all on 1928, they may be placed in a chain. This is done in line 9 by using the arithmetic average. These chain relatives are the index numbers for different years on the chain base method in respect to the year 1928. The process of chaining together the link relatives is as follows:

Average link relative for 1929 referred to 1928 is 98,
Average link relative for 1930 referred to 1929 is 75,
Average link relative for 1931 referred to 1930 is 73.

Then $\frac{98}{100} \times 75$ will give a chain relative index for 1930 on 1928,

and $\frac{98}{100} \times \frac{75}{100} \times 73$ „ „ „ 1931 on 1928.

Further chaining of link relatives has been done in similar manner.

The chain base method has two advantages. Firstly, it enables a direct comparison between one year and the year succeeding it. This is far more useful in business and commerce than the indirect comparison through a remote fixed base. Secondly, it makes possible the dropping of old items and inclusion of new ones, a necessity not infrequently felt when computing a series of index numbers over a period of time because of some commodities going out of use and new ones coming into fashion.

Reversibility of Index Numbers.

An important property, which an index number should possess, is its reversibility. That average which caters for reversibility of an index number based on it has a superior claim for use in the construction of index numbers than any other average which does not. Reversibility means that the index for the current year based upon the base year and the index for the base year based upon the current year should be reciprocal to

each other That is, the following equation should be satisfied

$$P_{01} \times P_{10} = 1\text{, or, } P_{01} = \frac{1}{P_{10}}$$

where P_{01} stands for index for the current year on the base year omitting the factor 100 (i.e., for price change in current year compared with base year), and P_{10} stands for index for the base year on the current year without the factor 100, (i.e., for price change in base year compared with current year)

The arithmetic average of relatives is not reversible.—Line 3 of table 31 below gives in column (d) the arithmetic average of relatives for commodities A and B for 1931 on 1930 as the base and in column (e) the arithmetic average of relatives for 1930 based on 1931 These are, respectively, 130 5 and 78 35, so that $P_{01} = \frac{130.5}{100} = 1\,305$ and $P_{10} = \frac{78\,35}{100} = 7835$

Now, $1\,305 \times 7835 = 1\,02$ which is greater than 1 Therefore the arithmetic average of relatives is not reversible

Table 31 *Testing the Reversibility of Index Numbers*

Line	Commodities	Price in 1930	Price in 1931	$\frac{\text{Year 1931}}{\text{Year 1930}} \times 100$	$\frac{\text{Year 1930}}{\text{Year 1931}} \times 100$
	(a)	(b)	(c)	(d)	(e)
		Rs.	Rs.		
1	A	10	15	150	66.7
2	B	45	50	111	90
3	Arithmetic average			130.5	78.35
4	Geometric mean			129	77.5

The Geometric mean of relatives is reversible.—Line 4 of table 31 gives in columns (d) and (e) the geometric average

of relatives. These are 129 and 77.5 approximately, so that $P_{01}=\frac{129}{100}=1.29$, and $P_{10}=\frac{77.5}{100}=.775$. Their product is 1 (allowing for the adjustment of decimals). Therefore, the geometric mean of relatives is reversible.

There is yet another way of looking at the reversibility of index numbers. If a relative shows an average increase of, say, 25 per cent. from the base year to the current year, then this should also be capable of being described as a decrease of 20 per cent. from the current year to the base year. In table 31, using the arithmetic average we find that the level of prices in 1931 is 30.5 per cent. higher over the prices of 1930. We might, therefore, say that the prices of 1930 are lower than those of 1931 by 23.4 per cent. of the latter; that is, if the index for 1931 on 1930 is 130.5, it should be (100—23.4)=76.6 for 1930 on 1931, but actually it is 78.35 as shown in column (e), line 3. Therefore, the index number based on arithmetic average of relatives does not possess the property of reversibility. Using the geometric mean we find that the level of prices in 1931 is 29 per cent. higher over the prices of 1930. We might, therefore, say that the prices of 1930 are lower than those of 1931 by 22.5 per cent. of the latter; that is, if the index for 1931 on 1930 is 129, it should be (100—22.5)=77.5 for 1930 on 1931, which actually is the case as shown in column (e), line 4, table 31.

It is clear, then, that geometric mean stands this test of efficiency and can be said to perform satisfactorily its function of showing the required change in the phenomenon under study. Consequently, geometric mean is more suitable than the arithmetic mean or the median. Geometric mean can also be used with the chain base method. It is used by the *Board of Trade* in England in the construction of wholesale price indices on the chain base principle. We have used the arith-

metic average in table 29 It is interesting to note that the geometric mean is used in the construction of Index numbers of wholesale prices of certain articles in India and of the Capital Index of Indian Industrial Activity

Base shifting

Not infrequently it becomes necessary to shift the base of a series of index numbers from one year to another For instance if a series of indices based on the year 1914 is to be compared with a similar series whose base year is 1939 then in the interest of valid comparison it will be necessary to convert the former series into a new one with 1939 as the base year That is the base of the former series will be shifted to a new one

From what has been said above in regard to reversibility of index numbers it is clear that index numbers based on geometric mean of relatives can be shifted from base to base without error by what may be called the short-cut method illustrated in the example accompanying table 31 But it is not possible to shift the base *without error* by using the same method when arithmetic average has been used in averaging the relatives This short-cut method is of course, not possible to apply when median is used in averaging the relatives.

In addition to this short-cut method another method for base shifting is also available *viz* re-computing the relatives of each individual item on the new base and averaging their total that is reconstructing the entire series This method should be used for shifting the base when arithmetic average and median have been employed in averaging the relatives of a series while both this and the short-cut, methods shall yield identical results when geometric mean has been used.

The System of Weighting.

The 'Unweighted' Index Number.

So far, in the construction of index numbers we have used simple averages and no *special* assumption has been made concerning weights. Distinction is very often made between weighted and unweighted index numbers, but it should be noted that every index number is weighted in some form. In computing the simple average of relatives each relative is counted once. Therefore, apparently, weights are unity in each case. A further thought would reveal that the change in the price of a commodity from one date to another is related to the commodity's price level on the first date. If in the base year the price of a commodity is unusually high, it will have an influence to correspond on the total, that is, it would have the same effect as actual weighting would. This can be easily verified by recalculating a given series of index numbers upon a few different bases by using the arithmetic average of relatives and then noticing that every fresh series differs, not only in the absolute values of the index numbers, which is immaterial, but also in the relative values of the indices, which is significant. That this would be so is evident from the fact that index numbers where simple arithmetic average is used are not reversible. We may, therefore, conclude that even the so-called 'unweighted' index numbers are arbitrarily or haphazardly weighted, the arbitrary element being exercised by the choice or shifting of base year. We may also say that when simple average of relatives is used change in the base year is equivalent to change in weights.

Implicit and Explicit Weighting.

In those index numbers which are termed 'weighted' the weights are chosen according to some systematic plan.

Weights may be implicit or explicit. Implicit weights relate to the selection of commodities themselves. If a particular commodity or a commodity of the same general class is included say 3 times in the list of prices, the weight given to the commodity is 3. For instance 3 varieties of sugar may be included. Varying emphasis is thus, given to the different items while selecting the commodities by the number of times a given commodity is included in the selection. Many of the so-called unweighted index numbers may in fact be indirectly or implicitly weighted. For instance the *Calcutta* and the *Bombay* Wholesale Price Index Numbers are implicitly weighted.

In assigning **explicit weights**, weights proportional to the relative importance of different items are used. But what considerations determine this relative importance? This enquiry is essential because weights should either be appropriate or they should not be used at all. Now, in constructing an index to show general changes in prices the weights assignable to wholesale prices may be several for instance the quantity of goods placed on the market value of goods produced values consumed and so on. Different systems of weighting would yield different results. The difficulty then is which of these or other similar criteria should be accepted as correct? This difficulty is not easy to solve. Therefore it appears that weights may better be ignored. This idea is strengthened by the fact that weighted results are almost identical with the unweighted ones if weights are chosen according to chance.

Nevertheless, the problem of selecting weights is one of practical concern. The merits of weighted and unweighted indices can be understood only by comparing them. If a properly weighted index agrees with an unweighted one weights may be dispensed with if it does not, weights ought to be used. According to Bowley[1] paucity of data might make the

[1] See Bowley A.L., *Elements of Statistics* 1920 ed., p. 206

inclusion of weights necessary and the popular desire for concrete measurements might make a fine show of weighting expedient. Weighting seems necessary also because of the heterogeneous character of the series from which indices are computed. Most wholesale price indices in the U.S.A. are weighted. Weighting is, indeed, essential in constructing cost of living and business activity indices, as we shall see later.

Methods of Weighting.

Two methods of explicit weighting may be distinguished: The Weighted Average of Relatives (Ratios) Method and the Weighted Aggregate of Actual Prices Method. The latter is known as the Aggregative Method also.

Weighted Average of Relatives.—According to this method *price relatives are weighted by values.* Values are obtained by multiplying quantities with their respective prices. The sum of the products of price relatives of the current year and values of the base year divided by the sum of the weights gives the weighted arithmetic average of relatives, which is the required index number for the current year. Symbolically,

$$\text{Index Number for Current Year} = \frac{\Sigma IV}{\Sigma V}$$

where, I stands for price relative and V for value (weight). Table 33 on page 253 demonstrates the working of this method. Weighted median and weighted geometric mean of the relative may also be computed.

Aggregative Method.—According to this method *prices themselves are weighted by quantities,* since total value is equal to price $\times$ quantity. The products of actual prices of the current year and quantities of the base year are summated. This sum is expressed as ratio in relation to a given base. This

ratio is the index number for the current year Symbolically

$$\text{Index Number for Current Year} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

where p_1 stands for price in current year

p_0 stands for price in base year

q_0 stands for quantity in base year

Table 32 on page 252 demonstrates the working of this method

In the above case the weights are fixed. If quantities for all the years for which it is desired to calculate the index numbers are available the weights may be made to vary from year to year quantities for different years being used as weights for their respective years Several formulae have been suggested for this purpose. We shall however confine ourselves to the Crossed Weight Formula given by Fisher which is supposed to be highly satisfactory

Fisher's "Ideal" Formula.

Professor Irving Fisher[*] after an elaborate examination of 134 possible formulae concluded that a scheme of cross weighting should be used and gave a Crossed Weight Formula which is also named as Fisher's Ideal Formula It is

$$\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q}{\Sigma p_0 q_1}}$$

This formula requires four sets of aggregates viz

(1) $\Sigma p_1 q_0$ Current year price × base year quantity

(2) $\Sigma p_1 q$ Current year price × current year quantity

(3) $\Sigma p_0 q_0$ Base year price × base year quantity

(4) $\Sigma p_0 q_1$ Base year price × current year quantity

[*] See Fisher Irving *Making of Index Numbers* 1922

The first aggregate is divided by the third, and the second by the fourth. The two resulting relatives are multiplied together and square root of the product is extracted. Fisher calls this formula as "ideal", since it neutralizes the types of bias which are found in measuring prices and quantities. The system of weighting has been so designed in the formula that the resulting index satisfies two basic tests, *viz.*, Time Reversal Test and Factor Reversal Test.

Time Reversal Test.—It has already been indicated in connection with the "Reversibility of Index Numbers"[4] what time reversal test implies. According to Fisher this test may be described as follows:

"The test is that the formula for calculating an index number should be such that it will give the same ratio between one point of comparison and the other, *no matter which of the two is taken as base.*

"Or, putting it another way, the index number reckoned forward should be the reciprocal of that reckoned backward."[5]

This implies that the following equation should be satisfied:

$$P_{01} \times Q_{10} =$$

This, again, means that if an index shows that between 1938 and 1942 prices doubled, then it should also show that the level of prices in 1938 was one-half of that in 1942 when measured from 1942.

Factor Reversal Test.—A second fundamental test by means of which good index numbers can be detected is the factor reversal test. Regarding this test Fisher says:

"Just as our formula should permit the interchange of the two times without giving inconsistent results, so it ought to

[4] See pages 237-239.
[5] Fisher, Irving, *Op. Cit.*, p. 64.

permit interchanging the prices and quantities without giving inconsistent results—i e the two results multiplied together should give the true value ratio *

This implies that the following equation should hold good

$$P_{01}\times Q_{01}=\frac{\Sigma p_1\,q_1}{\Sigma p_0\,q_0}$$

where P_{01} stands for the price change for the current year on the base year Q_{01} for the quantity change for the current year on the base year p_1q for the total value (price × quantity) in the current year p_0q_0 for the total value in the base year and $\frac{\Sigma p_1\,q_1}{\Sigma p_0\,q_0}$ for the ratio of the total value in the current year over the total value in the base year

Fisher s Ideal formula not only satisfies both the above tests but is also simple and easy to calculate from the practical point of view Therefore of the 134 possible formulæ which Fisher analyzed the Ideal is ideal But this formula requires statistics of quantities for the base year as well as the current year These statistics are generally not available for every current year They may be available at each successive census of production if such censuses are taken in a country Therefore the choice has to lie with the use of fixed weights i e quantities of the base year or the year supposed to be typical.

Summary and General Remarks.

The technique of construction of price index numbers may be summarised as follows —

(1) Select a reasonable number of representative commodities

* Fisher, Irving *Op Cit* p ~2.

(2) Arrange for obtaining their regular wholesale prices from

(i) either, standard trade journals,

(ii) or, leading dealers of representative centres.

(3) Average the price quotations, and obtain monthly or yearly average prices as the need may be.

(4) Reduce the average prices to percentages, *i.e.*, price relatives, on

(i) either, the fixed base method, where

(*a*) the fixed base may be a fixed year, or

(*b*) it may be an average of a period of years,

(ii) or, the chain base method.

(5) If the fixed base method is used, compute a simple average of relatives, using the arithmetic average, median or the geometric mean. Theoretically, the advantage lies with the geometric mean.
If the chain base method is used, chain together the link relatives.

(6) If weighting is necessary, compute

(i) either, the weighted average of relatives,

(ii) or, the weighted aggregate of actual prices.

Thus, we have discussed two important methods of constructing index numbers, *viz.*, the Average of Relatives Method and the Weighted Aggregate of Actual Prices Method. In the former method, the average may be Simple or Weighted.

A comparison of the unweighted index numbers calculated on the fixed base method in table 28 and of the unweighted index numbers calculated on the chain base method in table 29, and also of the weighted index numbers which can be calculated from the same figures would reveal that different methods yield

different results but **all index numbers—without any exception—point in the same direction** Therefore an index may be relied upon so far as the tendency shown by it is concerned without being trustworthy to the last digit It is not the absolute value of an index number that matters What matters is the general trend shown by it or by a series of index numbers

COST OF LIVING INDEX NUMBERS

The methods of weighting discussed above are more particularly used in the construction of cost of living index numbers These indices are designed to study the effect of changes in prices on the people as consumers or, in other words to study the average increase in the cost of maintaining the standard of living in a given year unchanged from that in the base year General index numbers fail to afford us an exact idea regarding the effect of the change in the general price level on the cost of living of different classes of people since a given change in the level of prices affects different classes of people differently Therefore to obtain a measure of the general movement of prices of those commodities which enter into the consumption of different classes of people Cost of Living Index Numbers are compiled

Difficulties in Constructing Cost of Living Indices.

Standard of living varies with income or occupation Therefore one single cost of living index will not be truly representative of people of different incomes Consequently index numbers should be compiled separately for different classes of people But standard of *living* also varies with region or place in which people reside This difficulty can be solved by compiling index numbers separately for different localities or different homogeneous zones Again same classes of people at the same time do not spend their income in exact

ly similar proportions on different objects. The best that can be done to obviate this difficulty is to collect a reasonable number of sufficiently accurate samples of family budgets from the same class of people to have a general idea of the proportions of expenditure on different objects by an *average* family. And, yet there is another difficulty that the same classes of people at different times spend their income in varying proportions. A change in the nature and quantity of commodities consumed may arise from a change in taste or fashions, or from an increased purchase of cheapening commodities and decreasing consumption of things becoming dearer. These factors, indeed, go a long way in explaining the change in the cost of living. But these changes cannot be taken stock of every year without incurring the huge expense of conducting fresh family budget enquiries. For this reason, it is assumed that the qualities and quantities of commodities consumed in the base year by a particular class of people remain the same for an indefinite period. These qualities and quantities, therefore, form the basis of the index number series. Another factor that causes a change in the standards of living is the change in the purchasing power of money. Cost of living index number confines itself to a measure of this factor alone. Further, people as consumers pay retail and not wholesale prices. Therefore, retail prices are taken into consideration in constructing cost of living indices. But retail prices vary from locality to locality. If cost of living index numbers are computed separately for different classes and different regions, this difficulty of variation in retail prices is also got over.

Construction of Cost of Living Index Numbers.

The first step, therefore, that is taken for the construction of cost of living index number series is to **decide the class of people**—industrial workers, clerks, etc.,—for which the index

numbers are to be compiled Next, a **sample budget enquiry** of the class concerned is made the sample covering a reasonably adequate number of families and conducted during a period reasonably free from abnormalities of very high or very low prices This budget enquiry would give precise information regarding (1) the nature qualities and quantities of commodities consumed by the people classified under the heads of food clothing rent lighting and fuel and miscellaneous groups (2) the retail prices of the different commodities (3) the proportion that the expenditure on each individual item of expenditure bears to the expenditure on the group to which it belongs (4) the proportion which expenditure on each group bears to the total expenditure This budget enquiry forms the basis of the index number series With it the **selection of commodities** whose retail prices are to be regularly obtained becomes easy It is important to note that a cost of living index number should include only those commodities under the food clothing etc groups which are generally used by the class of people concerned which are not subject to wide variations in quality nor to seasonal alterations in supply and for which regular and comparable quotations of prices are obtainable **Retail price** quotations should be obtained from the localities in which the class of people concerned reside or from which they usually make their purchases The **sources of price quotations** may be either standard trade journals or publications of government or municipalities or typical businessmen in the locality concerned From these regular price quotations **average prices** are calculated in the same manner as they are done in the case of general index numbers

To convert these average prices into index numbers the average prices or their relatives must of necessity be **weighted,** because the average consumer is not recompensed for a rise in the price of say cotton cloth by an equal fall in

that of cement. Different objects of his consumption have different importance in his budget. They must be assigned their relative importance. For this purpose one of the two systems of weighting may be applied: (1) The Aggregate Expenditure Method, and (2) The Family Budget Method.

Aggregate Expenditure Method.

This method is the same as the Weighted Aggregate of Actual Prices Method already discussed on pages 243-244. Table 32 demonstrates the calculation of cost of living index number for the artisan class in the Eastern U. P. by this method. (The figures in the table are imaginary). Quantities consumed in the base year have been taken as weights for the current year. The quantities consumed in the current year may be, and usually are, different from those consumed in the base year, but for the reason already indicated, *viz.*, the enormous cost involved in conducting a fresh budget enquiry every year, fixed weights, *i.e.*, quantities consumed in the base year, have been, and are, used as weights. This is also why Fisher's "ideal" method is difficult to follow in practice, for in using it quantities consumed in the current year should be known in addition to those consumed in the base year.

Table 32 *Construction of Cost of Living Index Number by the Aggregate Expenditure Method*

(1) Article	(2) Quantities consumed in Base year (1925)	(3) Unit	(4) Price in base year (1925)	(5) Price in current year (1941)	(6) Aggregate Expenditure in base year [cl. 2×cl. 4]	(7) Aggregate Expenditure in current year [cl. 2×cl. 5]
	q_0		p_0	p_1	$p_0\ q_0$	$p_1\ q_0$
			Rs.	Rs.	Rs.	Rs.
Rice	5 mds	per maund	6	8	30	40
Bajra Jowar	5 mds	,,	4	5	20	25
Wheat	1 md	,,	5	10	5	10
Gram	1 md		3	6	3	6
Arhar	5 md	,,	4	6	2	3
Other pulses	2 mds.	,,	3	4	6	8
Ghee	4 seers	per seer	1.25	2	5	8
Gur	2 mds.	per maund	2.50	[illegible]	5	10
Salt	12.5 seers	,,	4	5	1.25	1.6
Oil	24 seers	,,	[illegible]	25	[illegible]	15
Clothing	40 yds.	per yard	0.25	0.5	10	20
Firewood	10 mds	per maund	0.50	0.8	5	8
Kerosene	1 Tin	per tin	4	6	4	6
House-rent	—	per house	12	15	12	15
					$\Sigma p_0 q_0 = 120.25$	$\Sigma p_1 q_0 = 175.6$

$$\text{Index Number for the Current Year (1941)} = \frac{\Sigma p_1\ q_0}{\Sigma p_0\ q_0} \times 100$$

$$= \frac{175.6}{120.25} \times 100$$

$$= 146$$

Quantities consumed in any year (supposed to be typical) other than 1925 could also have been used as fixed weights. Similarly figures proportional to quantities consumed could also have been used in place of the actual figures.

Family Budget Method.

This method is the same as the Weighted Average of

Relatives Method already discussed on page 243. Table 33 demonstrates the calculation of cost of living index number for the same artisan class of the Eastern U. P. by this method, using the same data. **Values** consumed in the base year have been used as weights for the current year. For the reason already indicated values in the current year are not used as weights, and fixed weights are employed.

Table 33. *Construction of Cost of Living Index Number by the Family Budget Method.*

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Article	Unit	Price in base year (1925)	Price in current year (1941)	Price Relatives for current year	Weights Values consumed in Base year)	Product of Price relatives and weights [col. 5×col. 6].
		p_0	p_1	$p_1/p_0 \times 100 = I$	V	IV
		Rs.	Rs.		Rs.	
Rice	per maund	6	8	133.3	30	3999
Bajra, Jowar	"	4	5	125	20	2500
Wheat	"	5	10	200	5	1000
Gram	"	3	6	200	3	600
Arhar	"	4	6	150	2	300
Other pulses	"	3	4	133.3	6	799.80
Ghee	per seer	1.25	2	160	5	800
Gur	per maund	2.50	5	200	5	1000
Salt	"	4	5	125	1.25	156.25
Oil	"	20	25	125	12	1500
Clothing	per yard	0.25	0.5	200	10	2000
Firewood	per maund	0.50	0.8	160	5	800
Kerosene	per tin	4	7	150	4	600
House-rent	per house	12	15	125	12	1500
					$\Sigma V = 120.25$	$\Sigma IV = 17555.05$

$$\text{Index Number for the Current Year (1941)} = \frac{\Sigma IV}{\Sigma V}$$

$$= \frac{17555.05}{120.25}$$

$$= 146.$$

Values of any year other than 1925 could also have been used as weights Figures proportional to actual values could also be likewise used For instance instead of using 30 20 5 etc as weights we could have divided each of them by 5 and used 6 4 1 etc as wegihts

It will be seen that the cost of living index numbers by both the aggregative and the family budget methods exactly agree Indeed they should if the weights relate to the same year Family Budget method or Weighted Average of Relatives method is largely in use

In table 33 weighted average of all the articles has been *directly* computed This process can be improved upon by (i) dividing the articles into food clothing etc groups (ii) weighting their price relatives by the proportion which expenditure on each article bears to the total expenditure on the group (iii) obtaining weighted average index for each commodity group (iv) weighting the group index numbers by the proportion which expenditure on each group bears to the total expenditure and (v) obtaining the weighted average of the group index numbers The final weighted average is the required cost of living index number This process of double weighting is more scientific than that of computing a direct weighted average and is in general use In India cost of living index numbers are computed by this method as will be seen in the next chapter

Errors in Cost of Living Index Numbers.

The sources of error in cost of living index numbers lie in—

(1) demarcating one class of people from another incorrectly

(2) the faulty selection of representative articles entering into the cost of living of the class of people concerned

(3) the collection of price quotations—Information relating to clothing, for instance, relates, in part, to cloths rather than to clothes simply because fairly steady and reliable prices for the latter are not available;

(4) the faulty assignment of weights—Weights may be deliberately manipulated;

(5) the changes in demand of various articles or their prices in the period under investigation. For instance, the budget enquiry for the construction of cost of living index numbers for industrial labour at Kanpur extended over two decades *viz.*, 1914 to 1934 during which period the price of firewood changed from 6½ annas per maund in 1914 to 13 annas in 1919 and to nearly 8 annas in 1930.

Unsatisfactory Character of Cost of Living Index Numbers.

Even if errors of the types enumerated above are not allowed to enter into the construction of cost of living index number, the index is not fully dependable. The reason is simple. The total monthly expenditure of two families of the same class may be equal, but the distribution of the expenditure over different objects may considerably vary according to the number of persons in the family, their age, sex, religion, caste and mode of living, and also according to the rise and fall in the prices of articles consumed. Index number does not consider the variations in the expenditure of the individuals. It considers only average or normal cases. Therefore, an index number pointing to the change in the average cost of living cannot be applied to every individual case in the class. It should be taken only as a guide to the direction, the general trend, of the purchasing power of money for the class of people for whom it is meant.

Further in constructing cost of living index we proceed on the assumption that the quantities or the values consumed in the base year or the typical year do not change, whereas they actually do Standard of living of the same family undergoes change as time elapses and as prices, tastes etc, change. This fact is not taken into account by the index. The objection may be met by answering that the index number considers the increase in the cost of maintaining unchanged the base year standard of living True but is there a particular sanctity about the base-year standard of living? The base-year standard may not be adequate Improvement in it may be necessary Therefore to make the index truly representative, family budgets should be collected regularly after an interval of a few years and new weights adopted and commodities and their qualities and quantities modified in the light of every fresh inquiry for subsequent years

INDICES OF INDUSTRIAL ACTIVITY.

If it is desired to study the general change in the industrial activity in a country over a period of time, evidence of this change may be found in changes in the output of the various industries of the country The first step therefore, will be to collect information relating to the production of different groups of industries. Information may, for instance, be had for the following groups of industries

1. Mining—coal iron ore petroleum gold manganese
2. Metallurgical—steel work rolling mills foundries etc.
3. Mechanical—locomotives ship buildings railway rolling stock etc.
4. Textile—cotton woollen jute silk etc
5. Industries usually subject to excise duties—distilling of alcoholic beverages brewing sugar, matches and tobacco etc.

6. Other important industries—chemical, cement, glass-ware, flour-milling, oil-crushing etc.

As the statistics of production of these industries are received from year to year, those of the base year are put down at 100 and those of the subsequent years expressed as a percentage of the base year. These relatives are multiplied by weights assigned to them in proportion to the importance of the industries in the country. The weighted average— arithmetic or geometric—of the relatives gives the index number of industrial activity for the country.

INDICES OF BUSINESS CONDITIONS.

To attempt a study of the changes in the business conditions of a country, it would be necessary to collect far more comprehensive data than are required for computing indices of industrial activity. Professor Pigou selected the following series for a study of the changes in business conditions of England:

1. Unemployment percentage.
2. Consumpton of pig-iron.
3. Prices in England.
4. Rates of discount on three months' bills.
5. Volume of manufactured goods.
6. Agricultural production.
7. Yield per acre of nine principal crops.
8. Index of production from mines.
9. Clearings of London Clearing House.
10. Increase of bank credit.
11. Credits outstanding.
12. Annual increase in the aggregate money wage.
13. Rate of real wages.

14 General aggregate consumption
15 Proportion of reserve to liabilities of the Bank of England.

These quantities may be converted into relatives referred to a base year From these relatives a weighted average may be obtained. This weighted average shall be the Index Number of Business Conditions It will afford a general idea of the average change in the business conditions of the country and serve as Economic Barometer or Forecaster of changes in business conditions through periods of depression, recovery prosperity and crisis Business conditions are never stationary They do change but the change, it has been found by experience in industrial countries of the west, particularly the U S A, is not fortuitous These changes are also not regular and periodic Business in general passes through well-defined major and minor changes Accordingly, it is possible to study their order, to measure their present conditions and to give a forecast of likely position in the future Both in England and in the U S A interest in this subject is growing

Uses of Index Numbers.

Index numbers reflect the movement of some quantity to which they relate Their peculiar character is that they exhibit the relative rather than the absolute aspect of such movement Index numbers are not restricted to the price phenomenon alone Any phenomenon which is stretched over a period of time and expressed numerically may be presented through them. They may for instance, be designed to show changes in wages, values of exports and imports prices of securities production of certain manufactures, circulation of notes etc. Different index numbers serve different purposes

General price index numbers measure general changes in prices and through them the value of money If general prices, as indicated by the price index number for a certain

year, double, the purchasing power of money for that year would be halved. Prices can be brought down or controlled through either the regulation of the supply of money and credit, or the regulation of production, or both. In any case, index numbers will provide an apparatus to study the fluctuation of general prices and a standard for keeping them steady in the interest of consumer, trade and public finance.

General price index numbers make possible a study of the movements of prices in different countries and of the fact whether they are fairly stable.

Cost of living index numbers indicate through their movement whether real wages are rising or falling, money wages remaining unchanged. They can be used to grant bonus to employees to meet the increased cost of living. Claims of labour for increase in wages, if they turn upon rising cost of living, can be shifted on the basis of cost of living indices.

Indices of industrial activity can be utilized to study the progress of general industrialisation of a country and the effect of tariff on the development of particular industries. When an industrial plan is being implemented, such indices are of immense use in judging the results of the policy adopted.

Indices of business conditions measure the change in the general economic activity of a country and afford an approximate idea of the fluctuations in the real national income of the country. They can be made to forecast economic events.

Investment index numbers are of great help to those interested in the stock market. Indices of imports and exports give an idea of the fluctuations in the foreign trade of a country.

It should, however, be noted that index numbers that are good for one purpose may not be useful for another. For instance, general index number indicating a rise in general

price level is not a good guide for movement in cost of living or cost of living index numbers for industrial workers are of no use for the upper classes

EXERCISES

(1) What is an Index Number? Why is it constructed?

(2) Describe the important problems involved in the preparation of an Index Number

(3) What considerations would weigh with you while constructing a wholesale price index number in connection with the selection of commodities and the base year?

(4) Give a list of at least 30 representative commodities for India and of their representative places for obtaining quotations

(5) Explain the difference between quantity price and money price How will you utilize the former in constructing a price index number?

(6) Distinguish between the Fixed Base and the Chain Base Methods of constructing index numbers and discuss their relative merits

(7) Which average do you think is appropriate to use in averaging the price relatives to arrive at the final index number Give your arguments

(8) What is meant by reversibility of an index number? Which index numbers are reversible?

(B. Com Luck 1930)

(9) Index numbers are economic barometers Explain the statement and mention what precautions should be taken in making use of any published index numbers

Show with the help of an example how you would convert the index number from one base period to another

(B Com Agra 1940)

(10) Describe with illustration the weighted index number of wholesale prices and show its importance

(B Com Nagpur 194)

(11) What are the different methods of assigning weights to price index numbers and cost of living index numbers? Which of them is suitable for general use?

(12) What do you understand by Time Reversal Test and Factor Reversal Test?

(13) Explain Fisher's "Ideal" Method of weighting index numbers and state the difficulties that are to be faced in using it.

(14) What points should be taken into consideration in choosing the base and determining the weights in the preparation of the Cost of Living Index numbers?

(B. Com., Agra, 1943).

(15) How are the cost of living index numbers calculated? Explain the different methods used for assigning weights to different commodities.

(B. Com., Alld., 1933).

(16) It is desired to find the difference in the cost of living in the years 1939 and 1943 in the case of (*i*) clerks and (*ii*) industrial labourers in a big industrial town.

Explain fully the necessary procedure to be adopted.

(17) What are the main sources of errors in cost of living index numbers? How can these errors be avoided?

(B. Com., Alld., 1938).

(18) Explain the method of studying changes in the business conditions of any country during a given period of time.

(19) Explain the meaning of Economic Barometers. How is this Barometer constructed, and how far it is being used successfully in forecasting economic events?

(M.A., Alld., 1938).

(20) Explain the uses of Index Numbers. Describe the procedure followed in the preparation of general and cost of living index numbers.

(B. Com., Agra, 1942).

(21) Explain the use of Index Number with the help of the following table, which gives the average annual wholesale price of jute in Calcutta in rupees per bale of 400 lbs. for the period 1914 to 1930:—

Year		Rupees	Year		Rupees
1914	..	78	1922	..	88
1915	..	54	1923	..	78
1916	..	67	1924	..	76
1917	..	56	1925	..	112
1918	..	73	1926	..	99
1919	..	102	1927	..	76
1920	..	98	1928	..	75
1921	..	94	1929	..	71
	..		1930	..	50

(B. Com., Cal., 1937).

(22) Given the following data what index numbers would you use for purposes of comparison? Give reasons

Year	Rice		Wheat		Jowar	
	Price	Quantity	Price	Quantity	Price	Quantity
19[illegible]7	9 3	100	6.4	11	5 1	[illegible]
193[illegible]	4 5	90	3 7	10	2 7	3

Prices and quantities are given in arbitrary units

(M.A. Cal [illegible]37)

(23) Explain what is meant by the Factor Reversal Test. Construct with the help of data given below Fisher's Ideal index and show how it satisfies the factor reversal test

	Estimated total produce in thousand tons in district Saran		Harvest Price per maund in district Saran			
	1931 32	1932 33	1931 32		1932 33	
			Rs	a	Rs	a.
Winter Rice	71	26	3	8	3	2
Barley	107	83	2	0	1	14
Maize	62	48	2	9	1	12

(M.A. Patna 1942)

(24) Which average would you use in computing the Price Index Number from the following data for 1943 on the basis of 1942? Give your reasons

Commodity	Unit	1942		1943	
Wheat	per maund	Rs	8-8-0	Rs	17-0-0
Ghee	per maund	Rs	50-0-0	Rs	75-0-0
Firewood	per maund	Re	1-0-0	Rs	00-8-0
Sugar	per seer	Rs	0-9-0	Rs	00-4 6
Cloth	per yard	Rs	0-5-0	Rs	00-2-6

(Figures in the above table are arbitrary)

(25) What is Chain Base Method? Describe it in connection with the construction of index numbers from the following data.

		Commodity		(Index Numbers)		
Year		A	B	C	D	E
1938	..	98	78	82	96	96
1939	..	100	82	78	100	100
1940	..	112	82	78	102	104
1941	..	110	84	84	98	98
1942	..	110	84	85	98	100
1943	..	120	90	90	100	100

(26) Following are the group index numbers and the group weights of an average working class family's budget. Construct the cost of living index number by assigning the given weights.

Group		Index Number for January 1943	Weights
Food	..	152	48
Fuel and Lighting	..	110	6
Clothing	..	130	8
House rent	..	100	12
Miscellaneous	..	90	15

(27) The following table gives the average annual prices of a few commodities in Allahabad for the years 1930, 1931, and 1932. Calculate the General Index Number for Allahabad for 1931 and 1932 on the basis of the prices in the year 1930, using the arithmetic average, median and geometric mean. Compare the results with those obtained by using the chain base method.

Commodity	Unit	Average Annual Price								
		1930			1931			1932		
		Rs.	a.	p.	Rs.	a.	p.	Rs.	a.	p.
Wheat	Maund	5	8	0	5	0	0	4	12	0
Rice	Maund	7	4	0	7	0	0	6	14	0
Arhar	Maund	6	0	0	6	0	0	6	0	0
Sugar	Maund	13	0	0	13	4	0	12	8	0
Salt	Maund	4	0	0	3	15	0	3	14	0
Ghee	Maund	60	0	0	58	12	0	58	0	0
Oil, Kerosene	Tin	4	2	6	4	0	0	4	2	0
Cloth	Yard	0	10	0	0	8	0	0	7	6
Fuel	Maund	1	2	0	1	0	0	0	12	0
Milk	Maund	5	5	6	5	0	0	5	11	0

(28) Construct appropriate Index Numbers and discuss the fluctuations in the quantity and value of (a) Raw Cotton and (b) Raw Jute exported from India for the period 1930-31 to 1935-36 using the average of the period 1926-27 to 1929-30 as base

Year	Raw Cotton		Raw Jute	
	Quantity (Thousand Tons)	Value Rupees (Lakhs)	Quantity (Thousand Tons)	Value Rupees (Lakhs)
1926—27 to 1929—30 (average)	609	59.41	826	29.24
1930—31	701	46.33	670	12.88
1931—32	473	23.45	587	11.19
1932—33	365	20.37	563	9.73
1933—34	504	27.53	748	10.93
1934—35	623	34.95	752	10.87
1935—36	607	33.77	771	13.71

(M.A. Alld. 1942)

(29) Construct the cost of living index number for 1940 on the basis of 1939 from the following data using the Aggregate Expenditure Method and the Family Budget Method.

Article	Quantity consumed in (1939)		Unit Price in 1939 Rs. a.		Price in 1940 Rs. a.	
Rice	6 maunds	maund	5	12	6	0
Wheat	6 maunds		5	0	8	0
Gram	1 maund		6	0	9	0
Arhar pulse	6 maunds		8	0	10	0
Ghee	4 seers	seer	2	0	1	8
Sugar	1 maund	maund	20	0	15	0
Salt	12 seers		20	8	18	0
Oil	20 seers		4	0	4	12
Clothing	50 yards	yard	0	8	0	1
Firewood	12 maunds	maund	0	8	1	2
Kerosene	1 tin	tin	4	0	5	2
House-rent		house	10	12	12	12

(30) The following are 23 price relatives that are available for the construction of an index number of prices:—

48, 53, 61, 61, 64, 64, 70, 71, 73, 76, 78, 81, 85, 93, 94, 96, 96, 97, 101, 102, 139, 143 and 144.

Regarding these as a statistical group, calculate their mean, median and a measure of dispersion.

Will you select the mean or the median as the appropriate average for the index number in question? Give reasons for your selection.

(M.A., Cal., 1935).

(31) Index numbers seek to set aside the irregularity of individual instances and replace it by the regularity of the big numbers.—Comment.

(32) An enquiry into the budgets of the middle class families in a city in England gave the following information —

Expenses on	food	rent	clothing	fuel	misc.
	35%	15%	20%	10%	20%
Prices (1928)	£150	£30	£75	£25	£40
Prices (1929)	£145	£30	£65	£23	£45

What changes in cost of living figures of 1929 as compared with that of 1928 are seen?

(B. Com., Luck, 1944).

CHAPTER XIII

INDIAN AND FOREIGN INDEX NUMBERS

We have already referred to some of the index numbers available in India in Chapter VII. Here it is proposed to study some well known Indian, British and American index number.

INDIAN INDEX NUMBERS

Current Wholesale Price Index Numbers

The following Index Numbers are being regularly published in India in the *Monthly Survey of Business Conditions in India* —

1. Calcutta Wholesale Price Index Number
2. Bombay Wholesale Price Index Number
3. Kanpur Wholesale Price Index Number
4. Index Numbers of Weekly Wholesale Prices of Certain Articles in India

Of the above indices the most generally used are the first two and the last.

Calcutta Index Number—This index includes 72 items which are divided into 16 groups. Cereals group includes 8 items Pulses 6 Sugar 5 Tea 3 Other food articles 9 Oil seeds 3 Mustard oil 2 Raw Jute 3 Jute manufactures 4 Raw cotton 2 Cotton manufactures 7 Other textiles (Wool and Silk) 2 Hides and skins 3 Metals 6 Other raw and manufactured articles 8 and Building materials 1. The prices on which this index

The Karachi Wholesale Price Index Number based on 23 commodities compiled by the Commissioner of Lahore Sind with July 1914 as base has not been available since June 1942.

The Madras wholesale Price Index Number which was compiled with prices in week ending 21st August 1939 equal to 100 has also not been available since June 1943.

number is based are the wholesale prices prevailing at the end of the month under review in Calcutta, which were published before October 1939 in the *Indian Trade Journal* and are since that date published in the *Wholesale Prices of Certain Selected Articles at Various Stations in India.* A separate index is computed for each group. The index for any group is the simple arithmetic average of the price relatives of the articles comprised in the group, with July, 1914, as the base. Weighting is introduced within each group by including more than one quotation for some items within the group. Thus under 'cereals' four varieties of rice are taken, wrereas wheat, barley, maize and oats have only one quotation each. To compute the general index number, a simple arithmetic average is taken of all the individual price relatives included in the computation. The general index number may also be considered as the weighted average of the group indices, the weight in each case being equal to the number of items included in the group. The index is compiled and issued monthly by the Department of Commercial Intelligence and Statistics, Calcutta. It is published in the *Indian Trade Journal*, the *Monthly Survey* and the Calcutta journal, the *Capital.*

Bombay Index Number.—This index includes 40 items which are divided into 11 groups. Cereals group includes 7 items, Pulses 2, Sugar 2, Other food 3. These four groups constitute 'All food' articles. The remaining 7 groups consist of 'All Non-food' articles. Among them Oil-seeds group includes 4 items, Raw cotton 5, Cotton manufactures 3, Other textiles 2, Hides and skins 3, Metals 5, and Other raw and manufactured articles 4. The prices on which this index is based are the wholesale prices prevailing in Bombay. Its construction is similar to that of the Calcutta index. Like the latter, it is also *indirectly* or *implicitly*. weighted by taking, for instance, 2 varieties of silk, 3 of wheat, 5 of raw cotton. Its base is also

July 1914 The index is compiled and issued by the Labour Office Government of Bombay in the *Labour Gazette* Along with the General Index Number group index numbers and All food and All Non food index numbers are also published.

Economic Adviser's Index Number.—The Index Number of Weekly Wholesale Prices of Certain Articles in India, commonly called Economic Adviser s index is of sensitive type. It is based on 23 commodities which are divided into four groups. Weekly and monthly average index numbers for the 23 commodities and the four groups are published along with the general All Commodities index. The four groups are (1) Food and Tobacco (2) Other Agricultural Commodities (3) Raw Materials (4) Manufactured Articles Besides two more group indices—one of Primary Commodities and the other of Chief Articles of Export—are also issued. The prices on which these index numbers are based are all India wholesale prices Geometric mean is used in their construction. Their base is week ending 19th August 1939 They are compiled and issued by the Office of the Economic Adviser to the Government of India

Inadequacy of Calcutta, Bombay and Economic Adviser's Indices.—The following points should be kept in view while making use of the Calcutta and the Bombay indices of wholesale prices —

(1) The price quotations for each commodity in both the cases refer only to one day in the month Therefore the indices cannot be regarded as sufficiently representative of the average price level during the month This is particularly so in times of abnormal price movements

(2) Each of these indices relates to the price level in one particular market But the different markets in the country differ considerably among themselves with regard to the relative degrees of importance of the various articles For instance

in Calcutta, rice is given a weightage of 4, while wheat gets that of only 1, but in the north Indian markets the position will be reversed. Therefore, these indices are not conclusive in discussions relating to All-India problems.

Partly because of the above limitations of the Calcutta and the Bombay wholesale price indices, and, may be, partly because of these index numbers being much higher than the Economic Adviser's index for the same month, the tendency to use the Economic Adviser's All-Commodities index in discussions relating to economic problems of India is increasing. But, it should be remembered that this index number is not a 'general-purpose' index. As its name implies, it is based on 'wholesale prices of certain articles in India.' 'Certain articles' which it includes are not the only representative articles for this vast country, whose inland and foreign trades are of considerable dimensions*.

Therefore, the necessity of the compilation of an All-India general-purpose index, based on a reasonably adequate number of representative commodities, cannot be overemphasized. The commodities may be divided into Food and Non-Food groups, and the system now followed by the British *Board of Trade* for the construction of its index is worth adopting as the model. The use of proper weights, geometric mean and chain base method would place the index on modern and scientific lines. It is necessary to compile such an index number in view of the fact that the main uses of wholesale price indices are in relation to national economic problems and for the study of general tendencies. They are considered in relation to movements of currency, exchange, foreign wholesale prices, indices of production, wages, retail prices etc. These purposes cannot be served by the Calcutta and the Bombay indices which can be utilized, for reasons already indicated, only in local (and not national) economic problems.

---

* For a critical estimate of the Official (Economic Adviser's) Index see S. C. Chaudhri's article on pages 428-429 in *The Eastern Economist*, Vol. 4, No. 14, dated April 6, 1945.

Discontinued Wholesale and Retail Price Indices.

The compilation of the following index numbers included *in the Index Numbers of Indian Prices* (quinquennial with annual supplements) has been discontinued

1 Index Numbers of Prices for Exported and imported Articles
2 Index Numbers of Retail Prices of Food Grains
3 Weighted Index Numbers of Wholesale Prices

Indices of Prices for Exported and Imported Articles —These indices include separate indices for (i) 28 exported articles (ii) 11 imported articles and (iii) all articles The all articles index number is generally known as the All India Wholesale Price Index Number In using these indices it must be kept in view that they are the unweighted arithmetic averages of the price relatives of the various commodities worked out with 1873 a rather old year, as base Another defect of these indices arises from the introduction in the years following the base year, of a few commodities whose quotations were not included in the index in the base year as also from the replacement of older varieties of some commodities by new ones at intervals of varying lengths Largely because of these factors and also because of the fart that the list of articles had not been revised since 1889 these indices particularly the All India Wholesale Price Index Number, had outlived their utility The Bowley Robertson Committee was therefore not in favour of the continuance of the series They have been discontinued since August 1941

Index Numbers of Retail Prices of Food Grains— These indices are the unweighted averages of the price relatives of seven commodities *viz*, rice wheat *jowar*, *bajra* gram barley and *ragi* worked out with 1873 as base The prices used in the computation are those reported by Provincial Authorities. These quotations are based on information collected by officials

in the *tehsil* or *taluq* centres from enquiries in *bazaar* areas. But, as pointed out in Chapter VII[2], the collection of these prices is not done with the care it deserves, with the result that official figures have been much different from those supplied by the traders. Index numbers compiled from unreliable prices cannot be regarded as reliable. This fact must be borne in mind while making use of the indices of retail prices of food grains.

Weighted Index Number of Wholesale Prices.—This index number includes 37 commodities of which 14 are articles of food, 17 of raw produce and 6 of manufactures. The method of weighting adopted is to take a number of quotations equal to the weight in the case of each commodity. The base year is 1871, but the figures have been re-calculated by shifting the base to 1873 for purposes of comparison with the other series of index numbers. In using this index number, it must be remembered that the re-calculated figures are subject to a certain margin of error since the arithmetic mean does not, as pointed out in the last chapter, satisfy the time reversal test. It is not safe to rely on this index number as a guide to price movements in general for the additional reason that certain important articles like groundnuts, pig-iron and steel manufactures are not included in it.

Cost of Living Index Numbers.

Twenty-seven[3] working-class cost of living index numbers are being regularly published in the *Monthly Survey of Business Conditions in India*, in addition to provincial bulletins or gazettes.

---

[2] See page 78.

[3] Besides these, cost of living indices are also being compiled for Jalgaon in Bombay, for a few more towns in the C. P., for Government servants drawing upto Rs. 30 per month in Meerut and Gorakhpur and Secretariat peons in Lucknow in the U. P., and for Bangalore in Mysore State, but they are not published in the *Monthly Survey*.

Diversity in Scope and Construction.—These index numbers are compiled on different bases. The cost of living index number for Bombay is compiled on year ending June 1934 as the base that for Ahmedabad on year ending July, 1927 and that for Sholapur on year ending January, 1928. They are compiled by the Bombay Labour Office Indices for Nagpur and Jubbulpore are compiled by the Department of Industries of C P and Berar with August, 1939 as the base, and are published in a special bulletin every month Indices for Patna, Muzaffarpur Monghyr Jamshedpur, Jheria and Ranchi are compiled by the Department of Industries of Bihar and for Cuttack by that of Orissa with average cost of living for five years preceding 1914 as the base The base for the index number for working class cost of living for Madras is year ending June, 1936 for indices for Lahore, Sialkot Ludhiana Rohtak and Multan in the Punjab is 1931 35, for index number for Kanpur in the United Provinces is August 1939, and for indices for Vizagapatam Ellore, Bellary, Cuddalore, Coimbatore, Madura Trichinopoly and Calicut is year ending June, 1936. Thus, the base period of the various index numbers varies from the quinquennium ending 1914 as in the case of centres in Bihar and Orissa to as recent a base as August 1939

For obtaining 'weights' for these indices family budget inquiries have been made from time to time in some of the provinces Detailed and comprehensive studies have been made only in a few places such as Bombay, Ahmedabad Sholapur, Madras City, Nagpur and Jubbulpore The Kanpur index is based on the tabulations of 300 out of 1500 family budgets of mill workers that were collected in 1938 39 by the Labour Office of the U P The weights used in the compilation of the indices for the Punjab centres were derived from only 138 family budgets of workers getting Rs. 50 or less per month which were collected in connection with the

investigations of the Royal Commission on Indian Labour. The weights used in the construction of Bihar and Orissa indices do not rest on any adequate statistical basis.

Further, neither is there any uniformity in the various provinces regarding the **agency** employed for the collection of prices nor regarding the **frequency** with which the data are collected. In some centres prices are collected weekly, in others fortnightly, while in the Punjab centres they are recorded only on the last day of each month.

The **scope** of the indices also shows great variations. Almost all the indices are fairly comprehensive in regard to the Food Group, but they show much variations among themselves with regard to other groups. The Jheria index does not include the Fuel and Lighting group, the indices for centres in Bihar, Orissa and the Central Provinces do not include House-rent, while the Clothing Group is somewhat unsatisfactory in most of the indices, firstly because some indices include very few items of clothing and secondly because the obtaining of comparable price quotations is difficult. The Bombay and the Madras indices are fairly comprehensive in respect of the Miscellaneous Group, but the Bihar and Orissa ones completely ignore these items.

Thus, there is a great deal of diversity in the scope and method of construction of the above-noted cost of living index numbers as between province and province. The base periods differ widely in time as well as in length; the 'weights' have been obtained as a result of inquiries made in the neighbourhood of the basic period in each case so that the several series refer to widely differing standards of living; the agency of collection of prices and the frequency of quotations show lack of uniformity; and, some of the series ignore important items like house-rent and miscellaneous articles. For all these reasons,

the cost of living indices relating to the different Provinces are not directly comparable with one another

The index numbers for the centres in the Provinces to Bombay Madras and the Punjab are similar in construction. The items in the list of articles consumed by the working classes are grouped under five heads *viz* food fuel and lighting clothing house-rent and miscellaneous Separate indices are worked out for the individual groups The index for a group is the weighted average of the price relatives of the various items in the group the weight assigned to any item being the ratio which the expenditure on this item bears to the total expenditure on all items included in the group The group indices are combined into a general index in a like manner A detailed study of the construction of the cost of living index number for industrial workers in Bombay would clearly show the whole process involved

Bombay Working-Class Cost of Living Index.—This index was first published in 1921 and was based on the aggregate consumption method in the absence of any reliable weights to be given to different items The Bombay Labour Office conducted the first inquiry into working class family budgets in Bombay City between May 1921 and April 1922 and a second inquiry between May 193 and June 1933 to ascertain weights proportional to the relative expenditure on the different items consumed by an average Bombay workers family The results of the second inquiry have been used in the compilation of the revised index commodities have been made as comprehensive as possible and the miscellaneous group has been added. This index and the indices for Ahmedabad and Sholapur are published in the *Labour Gazette* issued by the Labour Office of the Government of Bombay

The items included in the revised index are divided into five main groups *viz* food fuel and lighting clothing

rent and miscellaneous. The **food group** includes 28 articles, which are: rice, *patni*, wheat, *jowari, bajri, turdal*, gram, raw sugar (*gur*), refined sugar, tea, four varieties of fish, mutton, milk, *ghee* salt, dry chillies, tamarind, turmeric, potatoes, onions, brinjals, pumpkins, cocoanut oil, sweet oil and ready made tea. The expenditure on other articles which can be included in the food group has been proportionally divided among items of like nature included in the food group; for example, the expenditure on refreshments has been added to expenditure on ready-made tea, and that on sweet-meats has been divided equally between sugar and milk. **Fuel and lighting group** includes charcoal, firewood, kerosene oil and matches. **Clothing group** includes *dhotis,* coating, shirting, cloth for trousers, *sarees*, and *khans*. The figure adopted for **house-rent** is the average rent per tenement obtained as a result of the 1932-33 family budget inquiry. **Miscellaneous group** includes barber (shave), washing soap, medicine, *supari, bidis*, travelling to and from native place and newspapers. Thus, this index number includes 46 articles.

The price quotations for almost all the articles, except clothing articles, four varieties of fish, brinjals and pumpkins, are collected weekly by the officers of the Labour Office from two shops in twelve different industrial areas. The prices of all the clothing articles except *khans* are obtained from four different cotton mills having retail shops in Bombay City. Prices of fish, brinjals and pumpkins are taken from the Municipal records.

The method[4] adopted for computation of the index number is very similar to that of the British Ministry of Labour.

---

[4] With effect from the index for the month ending 15th May, 1943 the method of compilation of the index number for the cereals sub-group has been readjusted consequent on the introduction of rationing at Bombay in May, 1943. Besides, the indices from June, 1943, onwards relate to the calendar month and not to the month ended the 15th as before.

Price quotations for the current year are *first* expressed as percentages of the prices for the base year These percentages are weighted by the percentages which expenditure on each item bears to the *total expenditure on the group* to which it belongs and the products are summated Sum of the products divided by 100 gives the weighted average index for each group *The group index numbers are again weighted by the* percentage distribution of the expenditure on each of the group and then divided by the sum of the weights *The resulting weighted average is the final index.* The percentages by which group index numbers—are weighted are those arrived at as a result of the 1932 33 inquiry except in the case of the *miscellaneous group whose weight is 14 and not 25* which it may have been in view of the fact that the sum of the weights (percentages) for the first four groups comes to 75 The figure 14 represents the percentage *which expenditure on the actual* items included in the miscellaneous group bears to the total expenditure of the average working-class family The percentages for the different groups are

Food	47
Fuel and lighting	7
Clothing	8
House-rent	13
Miscellaneous	14
Total	89

Government of India's Latest Schemes

Because of the unsatisfactory character of the retail price *index numbers included in the Index Numbers of Indian Prices* and of the existing cost of living indices compiled in various centres of India there is an evident case for constructing retail *price index series in a scientific manner* and for compiling a cost of living index series on a uniform basis The Rau

Court of Enquiry, which was appointed in August, 1940 under the Trade Disputes Act, 1929 to investigate into the dispute regarding Dearness Allowance on the G. I. P. Railway, made the following observations in para. 111 of its report:—

> "None of the cost of living index figures at present available are entirely satisfactory ....The first requisite for any satisfactory revision of the allowances that we have recommended is the preparation of up-to-date cost of living index figures for three distinct classes of areas, city, urban and rural ...... We would accordingly recommend that the question of preparing and maintaining such figures for the purposes of the Central Government be considered by the Government of India."

Acting on this suggestion the Government of India formulated a centrally controlled scheme for the preparation and maintenance of cost of living index numbers in selected centres, a brief outline of which was circulated to Provincial Governments in October 1941 to which most of them gave a very encouraging response. The Third Conference of Labour Ministers held at Delhi in January 1942 concluded that it was advisable to ensure uniformity of technique in the compilation of cost of living index numbers in the various provinces. This was followed by the appointment of the Director, Cost of Living Index Scheme, by the Government to compile cost of living indices in selected centres of India on a uniform basis.

The Government, feeling that during the war[5] period occasions might arise when reliable figures indicating the changes in the level of retail prices would be urgently required, decided to proceed concurrently with a scheme for the compilation of retail price indices for those centres for which cost of living indices would also be ultimately compiled.

[5] (September, 1939 to...........)

Owing to difficult es of orgnisation it has been decided tenta tively to select 15 rural centres being way side railway stations situated i1 different parts of the country including Indian State t rritory and attempt the compilation of their retail price indices

Thus the Government of India are proceeding with three d stinct schemes —

1 The Main Cost of I iving Index Number Scheme.
2 Retail Price Index Nun ber Scheme Urban Centres
3 Retail Pr ce Ind x Number Scheme Rural Centres

The Main Cost of Living Index Number Scheme.—50 ce res—48 from d fferent provinces of India (excluding those of 1 North West I rontier and Madras) and Ajmer and Delhi c be 1 selected for which it is proposed to compile cost of liv ng u dices Family budget inquiries are to be conducted in these centres by the Provincial Governments or Administrations concerned It is hoped that some 20 000 family budgets would be collected with a view to determine the necessary 'weights' The lists of items for the Retail Price Index Number Scheme have been so drawn up that if and when family budget inquiries in the selected centres are completed and weights ascertained it may be possible to proceed immediately with the compilation of the necessary cost of living index numbers by making use of the retail price data collected by then

Retail Price Index Number Scheme, Urban Centres.— The centres selected for this scheme are the same as those selected for the main cost of living index scheme Retail price indices are being compiled in the Office of the Director, Cost of Living Index Scheme from weekly quotations of retail prices received by the Office from 53 urban centres Only 16 of these centres have a common base *viz* January 1943 With the rest the base month varies from centre to centre Indices for some of these urban centres are being published in the *Indian Labour Gazett*

Retail Price Index Number Scheme, Rural Centres.—The 15 centres selected for this scheme are divided into three zones: Northern, Eastern and Southern. Investigations regarding the food and clothing habits of the poorer sections of the community at these centres have been completed. On their basis, the lists of articles for which prices are collected have been drawn up, and certain shops have been fixed in each centre for the collection of prices regularly every week on a day appointed for this purpose. The task of the collection of prices has been entrusted to the station masters of these railway stations and their work is regularly supervised by the Inspectors of Railway Labour within whose beat the stations lie The returns, after careful scrutiny and tabulation in the Office of the Director, Cost of Living Index Scheme, are utilized for the compilation of monthly index numbers for all these centres. The base for these indices is January, 1943. They are being published in the *Indian Labour Gazette*

Industrial Activity Index

The inherent difficulties in the construction of index number of industrial activity are well-known. The absence of an Economic Census in India adds seriously to these difficulties Agricultural production in this country fluctuates considerably with seasons and climatic conditions, and is, for this reason, not suited to short-term inquiry covering a month or a quarter of a year. Besides, statistics of agricultural production are incomplete and not fully reliable. Therefore, in spite of its out-weighing influence, agricultural production cannot be included as a constituent of the Production Index. Bowley-Robertson Committee in their report recommended that an industrial production index should not be combined with that of agricultural production. If it is combined, the weight to be given to it would be of so considerable

a magnitude that it will swell up the general index of Indian business activity very high

Even the statistics of *industrial production* are *not quite* sufficient in India. In spite of this *Capital* the well known weekly journal of Calcutta has been publishing every month an Index of Indian Industrial Activity since March 1935

'Capital' Index of Indian Industrial Activity.—This index is published monthly and 1935 is taken as the base year. The series selected and the weights assigned to each item for computing this index are —

Series Selected	*Weight*
Industrial Production—	
1 Cotton Manufactures	9
2 Jute Manufactures	6
3 Steel Ingots	5
4 Pig Iron	8
5 Cement	5
6 Paper	3
Mineral Production—Coal	7
Rail & River borne Trade	24
Financial Statistics—Cheque Clearances	20
Trade Foreign & Coastal—	
Exports	4
Imports	3
Shipping Foreign & Coastal—	
Tonnage entered	3
Tonnage cleared	3

Since March, 1941 Trade Foreign and Coastal and Shipping Foreign and Coastal have been left out. Instead Notes in Circulation (base April 1935 to March 1936) with weight 6 and Consumption of Electricity with weight 7 have been included. The weighted geometric mean forms the general index and seasonal fluctuation are eliminated by means of a twelve months moving average Index for

cement appeared up to 1937-38 and has since been discontinued with the remark 'figures not available'. A specimen of the construction of this index is given in table 17, Chapter X, page 161.

Statistics for the above series are taken from the monthly publications of the Department of Commercial Intelligence and Statistics and from Statistical Summary of the Reserve Bank of India. This index does not afford an idea of the activities of people living in rural areas. And, even so far as urban people are concerned it is not fully representative. It does not include the production of sugar, tea, hides and skins which are quite important in the Indian industrial structure to-day. However, in the absence of complete and adequate statistical data no better index could be compiled.[6]

BRITISH INDEX NUMBERS

Wholesale Price Index Numbers.

Three important wholesale price index numbers that are compiled and maintained in Great Britain are:—

(1) Board of Trade Index Number.
(2) *Economist* Index Number.
(3) *Statist* Index Number.

Board of Trade Index Number.—The present series relating to this index begins with January 1935 and replaces an older series dating from 1920, which had replaced a still older series designed before the last Great War. The total number of

[6] In addition to the index numbers discussed above indices of Physical Volume of Production, of Prices of Securities, and of Values of India's Exports and Imports are published in the *Monthly Survey of Business Conditions in India*. A new Food Index, with the week ending August 25, 1940, as the base has also been started since February, 1944. The *Capital* also compiles and publishes a monthly Stocks and Shares index.

commodities included is 200, and the total number of quotations is 258 the difference being due to the fact that in certain cases the average of more than one quotation is used to get a better representative figure The commodities include food articles materials of industry and semi manufactured goods and are arranged in 11 groups Quotations are based upon market values. The index is not weighted in the ordinary sense of the word but is indirectly weighted by using two or more quotations for articles of special importance Price relatives are calculated upon the chain base method Geometric mean of the 11 groups is extracted on a footing of equality The base year has been successively 1913 1924 and 1930 The index is published in the *Labour Gazette*

Economist Index Number—This index number was originally framed in 1864 and has been revised twice in 1911 and in 1928 In its present form it comprises 58 commodities with 1927 as the base year Formerly arithmetic average was used in its construction but now unweighted geometric mean is used Results are published monthly and fortnightly It is compiled by the Economist an important periodical of Great Britain.

Statist Index Number.—This index number is really a continuation of a series begun by the late Mr Augustus Saurbeck who used 44 commodities and selected as his base the average of the monthly wholesale market prices of these commodities in the period 1867 77 He weighted the index not directly, but indirectly by taking two or more quotations for articles of special importance This method has also been followed by the *Statist*, a periodical of Great Britain, which continued this index from 1912 The same base is being maintained even now In its present form it is based on the wholesale prices of 19 foodstuffs and 26 raw materials. These

45 commodities are arranged in 6 groups. This index is valuable where a continuous record of figures over a long period is required, since it is presented in almost the same form in which it originated and also since its compilers publish every year full details of its construction.

Cost of Living Index Number.

The most important index number is that compiled by the British Ministry of Labour.

Ministry of Labour's Cost of Living Index Number.—This index number is designed to measure the average increase in the cost of maintaining unchanged the pre-War standard of living of the working-classes. The foodstuffs included represent about 75 per cent. of working-class expenditure on food. Retail prices are obtained from over 5,000 retailers, distributed among over 500 towns and villages. The weights used are based on the average expenditure of 1944 urban working class families. This information was collected by the Board of Trade in 1904. Prices in July, 1914 are used as the base of the index. The use of weights relating to 1904 instead of 1914 is considered reasonable on the ground that no great change took place in the standard of living between 1904 and 1914.

The weighted average increase in the relative prices of foodstuffs is combined with similar figures showing changes in rents, clothing, fuel and light, and other items. The weights used are: food 7½, rent 2, clothing 1½, fuel and light 1, miscellaneous ½; total 12½. The final index along with the five group indices, is published monthly by the Ministry of Labour. No allowance has yet been made for any changes in the standard of living or for any economies or re-adjustments in consumption and expenditure since 1914.

Indices of Production

The two important indices are

1 London and Cambridge Economic Service Index of Physical Volume of Production

2 Board of Trade Index of Industrial Production

London and Cambridge Index.—This index includes *agriculture and manufacturing and extractive industries.* Changes in the physical volume of production indicate the extent to which the country's resources are being used in industry and also indicate the results in terms of consumable goods The index is calculated in two forms (1) an annual index and (2) a quarterly index *Information in the annual index is tabulated* under the following heads

Group I Agriculture
II Principal Minerals
III Iron and Steel Engineering & Ship building Trades
IV Non Ferrous Metal Trades
V Textile Trades
VI Food Drink and Tobacco Trades
VII Chemical and Allied Trades
VIII Paper Printing and Allied Trades
IX Leather Trades
X India rubber Trade.
XI Building and Contracting Trades

The quarterly index is compiled upon the same general principles as the annual index subject to the omission of certain information which is not available quarterly Since 1929 weights assigned have been proportional to the net output of industries obtained as the result of the Census of Production of 1924 The system of weighting adopted is therefore base year weighting

Board of Trade Index.—This index differs from the annual (but not the quarterly) index of the London and Cambridge Economic Service by the exclusion of agriculture. But, certain branches of industries not covered by the latter are included in this index. The industries are classified into groups comparable, so far as possible, with the grouping adopted for Census of Production of 1924, *viz.*—

(1) Mines and Quarries, (2) Iron and Steel and Manufactures thereof, (3) Non-ferrous Metals, (4) Engineering and Shipbuilding, (5) Textiles, (6) Chemical and Allied Trades, (7) Paper and Printing, (8) Leather, Boots and Shoes (9) Food, Drink and Tobacco, (10) Gas and Electricity.

The objective of this index is the *net output* of the various industries, *i.e.*, the excess of the value of the products over the value of the materials utilized in their manufacture. Agriculture is excluded from the inquiry because of the fluctuations in agricultural production with seasonal climatic conditions and the consequent unsuitability of such production for an inquiry covering less than a year. The method actually adopted is to compare the best available statistics measuring the volume of production in the current quarter with the corresponding figures for 1924. Weights are assigned in proportion to the net output for 1924.

Indices of Business Activity.

Literature on the subject of business barometers and business activity indices is voluminous. The London and Cambridge Economic Service issues a monthly bulletin of comparable statistics upon every imaginable branch of economics and finance, such as, prices and wages, output and internal activity, foreign exchanges, finance. The Board of Trade, the Bank of England, the Economic Advisory Council, among others, issue periodical tables and charts relating to general econo-

mic conditions The most important index designed particularly for a study of business conditions, is that of the *Economist*

"Economist" Index of Business Activity.—This is a monthly index and goes back to 1924 It was revised in July, 1936 and recalculated with 1935 as the base year Its object is to measure changes in the economic activity of United Kingdom in quantitative—not monetary—units That is, it is designed *to afford an approximate idea of fluctuations in the* real national income The component series of the index and their respective weights are —

(1) Employment 10 (2) Consumption of coal 4 (3) Industrial Consumption of Electricity 2 (4) Merchandise on Receipts 3 (7) Building Activity 2 (8) Iron and Steel *Receipts 3 (7) Building Activity 2, (8) Iron and Steel* available for Consumption 2, (9) Consumption of Cotton 1, (10) Import of Raw materials 2, (11) Export of British manufactures 3, (12) Shipping movements 2, (13) Metropolitan, Country and Provincial Bank clearings 4 (14) Town clearing 1

All the series excepting building activity are corrected for *seasonal fluctuations Weighted geometric mean of the con*stituent series gives the Business Activity Index

UNITED STATES' INDEX NUMBERS

Wholesale Price Index Numbers

Following are the well known wholesale price index numbers in the U S.A —

1 Bureau of Labour Statistics
2 Federal Reserve Board s
3 Dun's
4. *Annalist's*
5 Fisher's

Bureau of Labour Statistics, Index.—This index number was a weighted average of relatives upto 1913 based upon the average price of 1890-1899. Since 1914 this index is the weighted aggregate of actual prices, and the weights now assigned are the amounts of goods marketed in 1919. Prices of 450 commodities are regularly collected by the Bureau. These commodities are arranged in 9 groups. Monthly and annual indices for the commodity groups, separately and combined, and reduced to relatives on base year, 1913, are published in *The Monthly Labour Review,* and in *Wholesale Prices,* both issued by the Bureau of Labour Statistics, Washington.

Federal Reserve Board's Index.—This index of wholesale prices has been prepared since October 1918—the series being calculated back to 1913. The price quotations, commodities and the method of calculating the index are the same as those of the Bureau of Labour Statistics' index, except that commodities are grouped in three major classes: raw materials, producers' goods and consumers' goods. Monthly and annual indices appear in *The Federal Reserve Bulletin, Washington.*

Dun's Index.—This index number is based upon the wholesale prices of about 200 commodities obtained from the principal markets of the U.S.A. The commodities are grouped into 7 classes, and weights are given according to average annual *per capita* consumption. The weighted aggregate of actual prices yields the required index, which is published in *Dun's Review,* New York. It is an index issued by private organisation.

Annalist's Index.—This index is computed by the *Annalist,* a New York financial journal. It is based upon 25 food products. The quotations are taken from Chicago and New York markets and are selected, it is held, so as to be representative of a theoretical family budget. This index is a simple arithmetic average of relatives, with average price for 1890-99

as the base No explicit weighting is used Weekly monthly and yearly indices are published in the journal This index is also issued by non government organisation

Fisher's Index.—Professor Irving Fisher of Yale University publishes weekly through a syndicate of American newspapers an index number of wholesale prices and its reciprocal the purchasing power of dollar The series began in the first week of January 1923 The quotations are taken from *Dun's Review* It is a weighted aggregate of prices of 205 commodities the actual quantities of each commodity sold in 1919 being the weights for their respective commodities The year 1913 is used as the base This index is also compiled by private organisation

Cost of Living Index Number.

An important cost of living index number issued by the United States Government is that compiled by the United States Bureau of Labour Statistics

Bureau of Labour Statistics' Index of Cost of Living.—This index number has been published by the Bureau since 1918 although the data have been computed back to December 1914 The price quotations refer to commodities consumed by working-class families They are in some cases submitted by storekeepers and are collected in other cases by Bureau's field agents The groupings are (1) food (2) clothing (3) rent (4) fuel and light (5) furniture and furnishings and (6) miscellaneous items. The system of double weightings as in the Bombay Cost of Living Index Number in India is adopted. Weights are based upon the result of a study of more than 12 000 family budgets in 92 localities in the U S.A. The year 1913 is used as the base Changes in the cost of living for the country as a whole and for different cities are regularly pub-

lished in the Monthly Labour Review, U. S. Bureau of Labour Statistics.

Indices of Production.

Among these indices, those compiled by Stewart, King and Snyder are important. The Harvard Committee on Economic Research also prepares them.

Harvard Committee's Index of Physical Production.— This Index is a quantity index prepared, separately and combined, for agriculture, manufacture and mining. Annual amounts of production of different items in these groups are expressed as relatives of the production in 1909, the base year. Weighted geometric mean of the group index numbers gives the combined index. The indices for the groups and the combined index are issued as adjusted and unadjusted.

Indices of General Business Conditions.

While dealing with indices of business conditions in Chapter XII it was pointed out that business in general, and certain of its phenomena in particular, pass through well-defined major and minor changes, so that it is possible not only to measure their present conditions but also to forecast what the future trend is likely to be. This service is performed by the Harvard Committee on Economic Research through its "Index of General Business Conditions."

Harvard Index of General Business Conditions.—As a result of an elaborate study of the data relating to U. S. A. for the period 1903 to 1914 it was found that there was a sequence in the movements in the speculative, business, and money markets which could be statistically measured and graphically presented. Accordingly, three curves, A, B and C, for Speculation. Business and Money respectively are presented on a chart which generally shows that movements in Curve A precede

those in Curve B and those in Curve B precede those in Curve C This movement occurs with such a regularity of sequence that the curves afford a logical basis for scientific business forecasting

The following series was used in the chart covering the trial period, 1903 to 1914 —

Curve A—Speculation
New York Bank Clearings
Prices of Securities

Curve B—Business
Wholesale Commodity Prices
Bank Clearings Outside New York City
Pig iron Production

Curve C—Money
Interest Rate on Commercial Paper
Loans and Deposits of New York City Banks

The index was presented in the form of a chart The following series was used for the period 1919-1924 —

Curve A—Speculation
Bank debits
Industrial Stock Prices

Curve B—Business
Bank debits for 140 cities outside New York City
Cyclical Index of Commodity Prices

Curve C—Money
Rate on 4-6 months good Commercial Paper
Rate on 4-6 months prime Commercial Paper

These new curves, although based on different data, have similar function to perform

In addition to the above is the Forecasting Composite Line prepared by the *Brookmire Economic Service* designed to forecast stock and commodity prices

EXERCISES

(1) What is the objective of an index number? State briefly the relevant conditions for its construction illustrating your answer by reference to the Index Numbers of Prices published by the Government of India.

(B. Com., Bombay, 1936).

(2) Explain the construction of the Calcutta Index Number of Wholesale Prices, and examine the advisability of treating it as conclusive in discussions relating to all-India problem.

(B. Com., Alld., 1944).

(3) Describe any index number in use in India at present for measuring changes in the wholesale price level, and point out its shortcomings.

(M.A., Cal., 1937).

(4) Examine critcally the Calcutta wholesale price index prepared by the Director-General of Commercial Intelligence and Statistics. In what respects is the new index prepared by the Economic Adviser to the Government of India an improvement upon the Calcutta index?

(M.A., Patna, 1943).

(5) How will you construct a cost of living index number of an Indian middle class family?

(M.A., Alld., 1937).

(6) How would you measure the cost of living in the United Provinces for a series of years? What are the difficulties involved, and how may they be solved?

(B. Com., Alld., 1943).

(7) Describe carefully how you would proceed to construct the cost of living index numbers for the U.P. (for the benefit of industrial labour). Would you allot weights according to 'Fisher's Ideal Method' or Family Budget Method? Give reasons in support of your answer.

(M. Com., Alld., 1943).

(8) Explain clearly how the "Capital" Index of Business Activity in India is calculated. How far do you consider it representative?

(B. Com., Alld., 1940).

(9) What statistical material would you utilize for preparing an Index of Economic Activity in India? How would you collate your data?

(M Com, Luck, 1942)

(10) Name the important Wholesale price index numbers and Cost of living index numbers published in India, England and the U S.A., and explain the construction of at least one of each type in each of the three countries

(11) What is the function of index numbers of business conditions? Explain it with an illustration of an actual index number of business conditions published in the U S.A., or England

(12) Write brief explanatory notes on the following —

(1) Saurbecks Index Number, (2) The Annalist Index Number, (3) Board of Trade Index Number, (4) The Statist Index Number, (5) The Economist Index of Business Activity, (6) 'Capital Index of Indian Industrial Activity, and (7) Bombay Cost of Living Index Number

(13) How will you make an estimate of the 'dearness allowance' that may be proposed to be given to industrial labour in Kanpur due to rise in the cost of living since the outbreak of the present War?

(14) If you are required to study the changes in business conditions in India, on which problems will you collect the information from official and non-official sources?

(15) Why are the working class cost of living index numbers available at present in India not directly comparable with one another? What steps are the Government of India taking to compile a new series on a comparable basis?

(M Com, Alld 1944)

(16) What is the technique of the cost of living index numbers? What is the method of their compilation in India and for what places are they compiled?

(B Com, Nagpur, 1943)

(17) Explain in plain English what you understand by the term 'Index of cost of living'. What are the difficulties commonly met with in computing an index of cost of living of the labouring classes in this country? What are the uses of a working class cost of living index? Have cost of living indices been computed for any industrial areas in this country? Give a critical account of any such indices, that you may be aware of, with suggestions for their improvement.

(M.A., Agra, 1944).

(18) Most industrial establishments in India are paying a dearness allowance to their employees on account of the high cost of living. Examine carefully how far an automatic linking of dearness allowance to the cost of living index is a proper solution of the problem.

(M.A., Patna, 1944).

(19) Give the salient features of the All-India Wholesale Price Index Number of Certain Articles, issued by the Economic Adviser, and examine critically its representative character.

(M. Com., Alld., 1945)

(20) Explain the construction and utility of the Bombay Working Class Cost of Living Index Number, and construct the index for April, 1944, from the following data:

Groups	Weights proportional to total expenditure	Group index Nos. for April, 1944
Food	47	247
Fuel and lighting	7	293
Clothing	8	289
House-rent	13	100
Miscellaneous	14	236

(B. Com., Alld., 1945).

CHAPTER XIV

DIAGRAMMATIC REPRESENTATION

An important function of the Science of Statistics is to present complex and unwieldy data in a manner such that they would be readily intelligible Classification and tabulation constitute the first step towards the attainment of this objective but even tables containing as they do a number of figures do not enable one to grasp the whole data at a glance Computation of simple derivatives statistical averages and index numbers constitutes further step in the direction of condensing the tabulated data But still the condensed material is presented in numerical form Numbers are not interesting to all To many they are dull and confusing, and if their number is pretty large, it would be difficult to compare them and observe their differences A long list of death rates and birth rates to take an example, relating to a large number of towns in a country, or to different countries of the world would tire one s eye and confuse his mind. It would not be easy for him to note the differences in death rates and birth rates of different towns or countries as the case may be Therefore it is necessary to adopt a device which may present huge mass of quantitative data, or their condensed form in a way that is at once comparable and appealing both to the eye and the intellect. For this purpose, the method of visual aids which comprises of presenting statistical material in pictures geometric figures and curves has been devised.

Usefulness of Diagrams

Diagrams carry with them the merits of **attraction** and **effective impression** One may not like to devote even a

minute to the study of a page—a small page—containing a number of quantitative figures; and, even if he devotes time, numerical figures may go out of his mind soon after he has studied them. But the same person may not—in most cases, would not—like to take his eyes away from a picture relating even to the same topic to which the numerical data did. Nay, he might invite others to have a look at the picture. And, if the picture has really attracted him, it need not be said that it would leave an effective impression on him. This is based on human psychology, and a successful advertiser or propagandist always exploits this psychology of the people to win his mark. A manufacturer of soap bars advertised for a considerable time that his bar, having the same price as that of his competitor, was much heavier than his competitor's, but did not find any improvement in his sales. And, when he advertised in pictorial form—a balance containing his bar on one side and his competitor's on the other, the pan containing his touching the ground, while the other much above the ground, and the words "For the same price" beneath the picture—he found to his pleasant surprise that the demand for his bars increased so much that he had to extend his plant.

It follows from the above example that diagrams are not only attractive and impressive, but also have the merit of rendering the whole idea **readily intelligible.** A man, to take another example, who has never seen or handled more than a few hundreds of anything may not understand how large the city of Bombay would be if he is told that its population in 1941 is 1,490,000. But, if he is living in Nagpur and is told that Bombay is nearly five times of his own city in the size of population, he would, not doubt, *try* to understand what it means. The idea, however, shall be more easily and readily grasped if this fact is represented to him diagrammatically—

e.g., an area may be divided into five equal parts one of which may be shaded and named Nagpur while the whole of the area would show the population of Bombay. It would be clear at a glance that Nagpur is only one fifth of Bombay.

Another merit of diagrams is the ease with which they make **comparison** possible. Population of Nagpur with Bombay's or the weight of one bar with that of the other in the above examples can be quite easily and readily compared.

Yet another characteristic feature of diagrams is that they **save much valuable time**, which would otherwise be lost in grasping the significance of numerical data.

Lastly, a chief merit of diagrams and graphs is that the entire data which when expressed in numerical form may be unwieldy and require a number of pages to write down are made **visible at a glance**.

For these merits of theirs diagrams are very useful in economic and social studies. A purely theoretical economist finds in them the basis for logical reasoning and easily explaining an economic law, such as the law of substitution or of diminishing utility. A practical economist may make his ideas impressive through diagrammatic representation. Knowing that the expenditure of the eleven provincial governments in India in 1940-41 on Industries totalled 115 lakhs of rupees and that on Police 1,120 lakhs of rupees he would do well to represent his idea diagrammatically rather than quoting the figures. When a social reformer is addressing an audience mere reading out of figures would make the hearing dull, tedious and tiring. But if he appears on the platform with pictures, diagrams and graphs his talk would be interesting, lively and impressive. A businessman or an administrator has hardly any time to devote to the study of a huge mass of figures however well arranged. But if he is presented with graphs showing the rise and fall of a certain activity or with

pictures and diagrams, it will hardly take him more than a few minutes to grasp the significance of the whole. It is, thus, evident that diagrams, charts, pictures, graphs and similar other visual aids serve a more useful purpose than any other device.

But, diagrams can be as much misused as they are useful. In advertisements and political propaganda they are often deliberately misleading, though they might be literally correct. The true statistician has to guard himself against mis-representation. Hence some general directions for drawing diagrams.

Directions for Drawing Diagrams.

It should be remembered that diagrams do not add anything to the meaning of statistics. They afford only a method of presenting them. However, when drawn and studied intelligently they bring to light the features of statistical groups and series; they show the various components of a group in relation to each other and to the group as a whole; they show the unity that underlies the scattered figures. They are, therefore, only a means to an end, the end being to make comparisons. Consequently, if there is only one isolated numerical quantity, there is no sense in presenting it diagrammatically. Similarly, if there are many figures, in no way related to one another and, therefore, having no common characteristic, they are incomparable and, therefore, need not be diagrammatically presented. For example, if we know that the monthly expenditure of a certain student is Rs. 50, his age is 22 years, the length of his nose is 1.325 inches and he has 20 books, we cannot represent Rs. 50, 22 years, 1.325 inches and 20 books by any kind of diagram, since the four numbers are incomparable. On the other hand, if we know that one student is 60 inches long and another only 48 inches, we can compare the two measurements and, therefore, represent

them diagrammatically It is then established that the method of diagrammatic representation can be made use of when there are at least two numbers which are similar in nature and character at least in one important respect and which also vary independently of each other

Another point that should be kept in view is that diagrams are not the substitutes for the *real* magnitude of the quantity they represent The size of a diagram changes with the change in the scale to which it is drawn The same quantity drawn to two different scales will yield diagrams of different sizes.

It is evident from the above that in the technique relating to drawing of diagrams the selection of the proper scale occupies an important place No rigid rules can be laid down for the selection of a proper scale but a general direction that can be given for the purpose is that that scale is the most suitable the diagram drawn to which would be neither so big as not to be visible at a glance nor so small as to look clumsy and indistinct and cover only a very small part of the space available The scale should be so chosen that the size of the resulting diagram would show the significant features of the numerical quantities for which it stands All principal details must be clear The vertical scale should be marked at equal unit spaces and the measurement of each unit space put down Generally the vertical scale should be shown on the left hand side of the diagram The horizontal scale should be given at the bottom of the diagram On each side the vertical and the horizontal the variable represented should be indicated. For instance the vertical line might show the amount in rupees and the horizontal the different countries The diagram should be neatly drawn with the help of drawing instruments It should be given a suitable heading The data represented diagrammatically should be given on a page adjacent to the one on which the diagram is drawn. If these

data are indicated in the diagram itself, care should be taken to see that the quantities are so placed in the diagram that they do not distort the visual impression conveyed by the diagram. To make distinctions clear, various kinds of dotting, lining, crossing, cross-hatching, or colouring should be used.

The drawing of diagrams, as a matter of fact, is not so difficult as the selection of suitable types of diagrammatic forms to depict a concise picture of the statistical data in hand. In selecting the most suitable diagram from among the varied forms of diagrams, the criterion should be that the diagrams selected should lead most *quickly* and with the greatest accuracy to the real meaning of the quantitative data. The test of a successful selection lies in the speed with which the quantities can be accurately studied with the help of the diagram selected.

Different Forms of Diagram.

Diagrammatic representation can be made in any one of the following ways:

(1) One dimensional diagram, *e.g.*, lines or bars drawn to a common scale.

(2) Two dimensional diagrams, *e.g.*, squares, and rectangles whose areas are made proportional to the given figures.

(3) Circular and angular diagrams, *e.g.*, circles whose areas are made proportional to given magnitudes, and which may be divided into sectors whose common unit is the degree.

(4) Three dimensional diagrams, *e.g.*, cubes, cylinders, blocks whose volumes are made proportional to the given figures.

(5) Pictograms, *e.g.*, statistical maps and pictures.

Technically there is no objection to using squares, rectangles cubes circles and pictures, but in practice, lines, bars and angular diagrams are the easiest to draw They can also be made sufficiently accurate Therefore, so far as possible they should be preferred The terms diagram, chart and graph are very often used without distinction, the same figure being given any of these names. We shall use the term diagram for the various forms pointed out above, and the term graph for curves only which would be dealt with in later chapters

One Dimensional Diagrams—Simple Bar.

A bar is merely a thick line whose width, though shown in the diagram is not taken into consideration It is shown merely to make the diagram look attractive. Therefore, these diagrams in which only one dimension is considered are called one dimensional diagrams Now, a bar may be shown as a simple bar or it may be divided into parts That one dimensional diagram in which the bars are not sub-divided, that is, where the bars represent the whole of a magnitude and not components of it, is called a **simple bar diagram.** We consider it first

To draw a bar diagram the height of the biggest bar should be adjusted to the size of the diagram. Some margin should be left all round the diagram to write down the title and the designation of units and scale The width of the bars should be neither too big nor too small. If the number of items is very large and the space very limited, thick bars may be replaced by thin lines as done in Figure 22 The bars should be drawn to a common horizontal or vertical base line, generally the horizontal base is used on which bars are made to stand vertically Horizontal base is used for the simple reason that comparison of one bar with another can

better be made in terms of height. In a single study all the bars must be of the same uniform width, separated by equal intervening spaces. Bars may be coloured, lined or dotted, but the colour used or lining or dotting done should be the same in all the bars in a single study. If bars are made to touch each other, that is, no intervening blank spaces are left, the diagram would look a continuous and blurred one with its top disfigured. Bar diagrams are not suitable for presenting continuous series such as that spread over a period of time. For this purpose graphical methods of presentation are used. Bar diagrams are suitable for representing spatial series.

Table 34 gives the yield of certain food crops in India for 1936-37,[1] which are diagrammatically represented in Figure 3.

Table 34. *Yield of certain food crops in India (including minor states)*—1936-37.

Crop	Wheat	Sugar Cane	Jowar	Gram	Barley	Maize
Yield (1,000 Tons)	8,513	6,289	5,401	3,817	2,311	1,836

The highest yield to be represented is 8,513,000 tons, A suitable scale has been selected to represent this yield properly, and the yield of other crops has been reduced to this scale. The lower ends of all the bars have been placed on the common horizontal base, so that comparison can be made between yields of different crops by comparing their heights. To make comparison easy, the bars have been arranged in descending order; they could have been arranged in ascending order as well. The scale is put down on the left-hand side, a little away from the biggest bar. Names of the crops are given below each bar.

[1] Thomas and Sastry, *Indian Agricultural Statistics*, page 119.

A suitable heading is given at the top The bars are separated by equal blank inter spaces

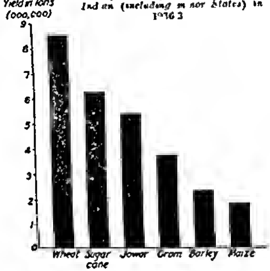


Fig 3

In interpreting Figure 3 it should not be said that in 1936-37 the yield of wheat in India was the largest among *all* the crops for there might be some crop other than the crops shown in the diagram whose yield may have been higher Actually, the yield of rice in the same year was 27 43 000 tons It can of course be said without any mistake that among the six crops represented in the diagram the yield of wheat was the highest and that of maize the lowest in 1936-37

In Figure 22 heights of 55 boys have been shown by vertical lines the method of construction being nearly the same as that of Figure 3

If instead of yield of food crops in India we were given the imports death rates or income *per capita* of different countries

of the world the above method of simple bars would have been followed to represent them diagrammatically. Again, if we were given imports and exports of different countries and it was desired to compare imports of different countries, exports of different countries and imports and exports of the same country, we could have extended the principle of simple bar diagram for this purpose as well. Choosing the horizontal or the vertical line as the base we could have placed two bars, one representing imports and the other exports of the same country, and separated this pair by an intervening blank space from another similar pair for another country, and so on, colouring the bars showing imports with one colour and those showing exports with another colour. This method is somewhat complex for simple comparisons, and can be replaced by that of sub-divided single bars.

One Dimensional Diagrams—Sub-divided Bar.

If a given magnitude can be broken up into the parts of which it is composed, or, if there are independent quantities constituting the sub-divisions of a total, bars sub-divided in the ratio of the different components may be used to show the *relationship of the parts to the whole*. For instance, if death rates and birth rates of different countries are given, bars may first be drawn to represent the birth, and from these bars portions from the bottom of the bars may be cut out in proportion to the death rates and coloured black to distinguish from the remaining white portion which would show the survival rate. Again, if the population of a number of countries is given, bars representing the population of different countries may be further sub-divided into two parts in proportion to males and females, one portion being shaded black to distinguish from the other. Technically, in these two examples we would say that the bar showing death rates has been

super imposed on that showing birth rate, or bars showing males and females have been super imposed on the bar showing the total population

Table 35 gives the value of exports and imports of India in total merchandise for three years These exports and imports may be added up and bars proportionate to the totals may be drawn the three bars being of unequal height because of the inequality of the total sea borne trade These bars may then be sub-divided into two portions, the lower one in each one of the three bars showing the exports and painted black and the upper portion remaining white, showing the imports Thus three comparisons will be possible at a glance —viz those relating to exports, imports and total trade in different years

Another method may be to draw three bars of equal length to show the total foreign trade which in each case may be put down at hundred, these bars should then be sub-divided according to the percentage which exports and imports in each year bear to the total foreign trade of that year

Let us suppose that the three bars are 5 inches in length each The total foreign trade for 1923 24 is about Rs 600 crores and the exports and imports are respectively Rs 363 crores, and Rs 237 crores so that they are respectively 60 per cent and 40 per cent of the total foreign trade The bar of 5 inches for 1923 24 would therefore, be divided into the proportion of 3 2 to represent 60 per cent and 40 per cent. respectively This single bar now represents percentange values of the imports and exports to total foreign trade of India separately The difference between this method and the former method should be carefully noted. The former method makes possible the comparison of actual values of imports with exports and of imports or exports with the total foreign trade, while the

latter method makes possible the same comparisons in percentage values.

If, however, the aim is to show the balance of trade the method of sub-divided bar diagram can be applied to subdividing either the bar for exports or for imports, whichever is greater into two portions, the one representing the imports or exports whichever is less and the other showing the balance of trade, positive or negative, as the case may be.

Table 35. *Value of Sea-borne Trade of India in Total Merchandise (including Govt Stores).*

Year	Imports	Exports	Balance of Trade
	Crores of Rs.	Crores of Rs.	Crores of Rs.
1921-2	282.59	248.65	— 33.94
1922-3	246.19	316.07	+ 69.88
1923-4	237.18	363.37	+126.19

To represent the figures of exports imports and balance of trade given in table 35 a common horizontal base line is chosen in Figure 4. For 1922-23 and 1923-24 two bars with their heights in proportion to the respective amount of exports are drawn; then, from the bottom of the two bars heights in proportion to imports are cut off and painted black. The heights of the full bars represent the values of exports, coloured portions represent the values of imports, and the remaining white portions show the favourable balance of trade indicated by plus sign in column 4, table 35. For the year 1921-22 the imports are greater than the exports. Therefore, first a bar representing the value of imports for the year is drawn on the same vertical

scale From the bottom of this bar a height equalling exports is cut off and left blank the small portion at the top being coloured black The height of the full bar indicates total imports white portion in it represents exports and the coloured portion at the top shows unfavourable balance of trade Were however the exports in a particular year exactly equal to the imports the bar would be painted black without the necessity of being sub divided The height of the bar would indicate total imports as well as total exports meaning that the balance of trade was nil Figure 4 makes possible the comparison of actual values Comparison of percentage values can also be made by drawing the bars on percentage basis in the manner indicated above

Value of sea borne trade of India in total merchandise (including government stores)

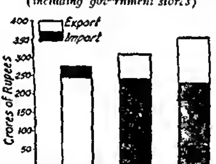


Fig 4

Note—*The shaded portion at the top of the bar for* 1921 22 *indicates unfavourable balance of trade while the white portions at the top of the bars for* 1922 23 *and* 1923 24 *indicate favourable balance of trade*

A single bar may be sub-divided into more than two sub-divisions as well Table 36 gives the proceeds cost and

profit and loss per table during three years. In the year 1938 there is a gain of Re. 1; in 1939 there is neither gain nor

Table 36. *Proceeds, Cost, Profits or Loss, per table during 1938, 1939 and 1940.*

Particulars	1938		1939		1940	
	Rs.	%	Rs.	%	Rs.	%
Proceeds per table	10	100	15	100	20	100
Cost per table—						
Wages	4.5	45	7.5	50	10.5	52.5
Other Costs	3.0	30	5.1	34	7.0	35.0
Polishing	1.5	15	2.4	16	3.5	17.5
Total Cost	9.0	90	15	100	21	105.0
Profit (+) or loss (—) per table	+1	+10	..	..	—1	—5

loss; while, in 1940 there is a loss of Re. 1. It is desired to represent the given data by sub-divided bar diagram drawn on percentage basis.

In table 36 the percentages of wages, other costs and polishing to proceeds per table are shown for the different years. In 1938 the profit is 10% of the proceeds and in 1940 the loss is only 5% of the proceeds, although the actual profit and loss in both the cases are the same, *viz*, one rupee. Sub-divided bar diagram drawn on the percentage basis would, therefore, be better in so far as comparisons of relative values are concerned; but it would not be suitable if

actual values are to be compared Figure 5 shows the diagram on percentage basis

Percentages of cost of and profit and loss on a table in 1938 1939 and 1940

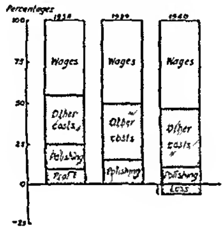


Fig 5

The construction of the diagram to represent the data in table 36 should be carefully studied First three bars to represent the proceeds per table are drawn and made equal to 100 The proportions of wages other costs and polishing are then cut down from the bars such that the same order for each of them is maintained in the bars The surplus in the bar for 1938 just above the horizontal base line indicates profit there is no surplus in the bar for 1939 In the bar for 1940 the deficit of 5% has been made good by extending the bar below the horizontal line in minus direction the

portion below the horizontal line showing the loss. Comparison in terms of percentages can be easily made from the figure. Bars could also be similarly drawn to compare actual values.

Although bars may be used for showing the sub-divisions of a large number of totals, it is not advisable to adopt the bar method for comparison if there are more than three or four components of each total, because in that case even if the same order is followed in the sub-divisions in each bar, the disparity among the figures may place them wide apart so that one type of component would not be opposite the other in all cases, and therefore it may not be possible to make comparisons at a glance.

Two Dimensional Diagrams—Rectangles.

The breadth of the bars, though shown in the diagrams, was so fat left out of consideration. It would be utilised now in drawing rectangular diagrams. A rectangle has two dimensions. Hence, its area, and not the height alone as in the case of one dimensional diagram, is taken to represent a magnitude. If several magnitudes are given, they may be represented by separate rectangles whose bases would be equal but heights would be proportional to the given magnitudes, so that their areas would stand in the same ratio as the given magnitudes. Rectangles are suitable for use in cases where two or more quantities are to be compared and each quantity is sub-divided into several components.

Table 37 gives the monthly incomes of two families and their expenses on different items. The data would be properly represented by rectangular diagram. The incomes of the families are Rs. 80 and Rs. 40. Therefore, in Figure 6, on the same horizontal line two rectangles are erected such that the

Table 37 *Family budgets of two families*

Item of expenditure	Family A, Income Rs 80: Actual expenses	Per cent age	Cumulative percentages	Family B, Income Rs 40: Actual expenses	Per cent age	Cumulative percentages
	Rs			Rs		
1 Food	32	40	40	20	50	50
2 Clothing	20	25	65	8	20	70
3 Shelter	8	10	75	4	10	80
4 Fuel and light	4	5	80	2	5	85
5 Miscellaneous	16	20	100	6	15	100
Total	80	100		40	100	

Percentage of Income spent by two families on different items of expenditure

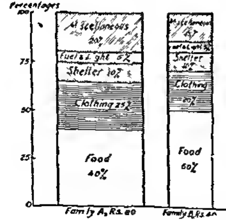


Fig 6

widths of the two are in the ratio of 80 40 or 2 1, and the

heights are equal. Thus the areas of the two rectangles are in the same proportion as the incomes. Each of the rectangles is, then, sub-divided into components according to the percentage of income spent on different items of expenditure. Percentages are shown in table 37. In order that the marking out of sub-divisions may be easy and convenient, cumulative percentages have been calculated. They are shown in the table in columns 4 and 7. The vertical scale is given on the left, and with its help cumulative percentages have been marked in the rectangles. Thus, for family A the first mark is put at 40, the second at 65, the third at 75, the fourth at 80, so that the remaining 20 is he percentage expenditure on miscellaneous items. Cumulation of percentages removes the chances of error in marking the sub-divisions.

The areas of the components of the rectangles are proportional to the actual expenses on various items, and the areas of the rectangles are in proportion to the income. Therefore, comparison of percentages of income spent over different items in the same family and in the two families is rendered easy. A glance at the heights of the rectangles relating to percentage expenditure on shelter and fuel and light would show that these heights are identical in the two. It implies that the percentage of income spent on these items in the two families is equal but the actual expenditure is different. The actual expenses stand in the same ratio as the areas of the components. Similarly, there might be a case in which the heights of the components may be different but their areas equal, implying that the actual expenses are identical but percentage expenditure is not, the percentage expenditure being in the same ratio as the heights of the components.

Rectangles are not used to represent only the family budgets. They are two dimensional diagrams and can, therefore, be also employed to show three different factors.

Consider table 38. It gives (1) price of a commodity, (2) its quantity sold and (3) different expenses of production and net profit. These facts have been diagrammatically represented in Figure 7, a close study of which would reveal that the heights of the two rectangles are in proportion to

Table 38. *Details of price, cost and quantity sold of two commodities.*

	I	II
Price of a Commodity ..	Rs. 2 per unit	Rs. 3 per unit
Quantity Sold ..	40	20
Value of raw-materials ..	Rs. 26	Rs. 24
Other expenses on production	Rs. 32	Rs. 21
Profits ..	Rs. 22	Rs. 15

the quantities sold, and the widths are in proportion to cost per unit, so that the areas of the rectangles are in proportion to the total proceeds in each case. These areas are sub-divided into their components, *viz.*, expenses on raw materials and production, and profit.

Details of cost of two commodities.

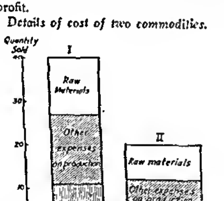


Fig. 7

It should be noted that rectangles are the only two dimensional diagrams which are capable of showing three different factors. Squares and circles which are also surface diagrams do not enjoy this property. Hence, the place of rectangles in diagrammatic representation is very important.

Two Dimensional Diagrams—Squares.

When it is desired to compare quantities which bear the ratio of 1:100, bar diagrams fail to answer the purpose, since howsoever small the scale selected be, the height of one bar will have to be made 100 times as great as that of the other with the result that one bar will be too tall and the other too small. In such cases bars are replaced by squares.

The side of a square varies as the square-root of its area. If, therefore, two figures 100 and 10,000 are to be represented by squares their sides would be taken in the proportion of $\sqrt{100}$: $\sqrt{10.000}$ or 1: 10 and not in that of 100: 10,000.

Table 39 gives the production of cane-sugar in four countries or 1938-39, arranged in descending order. Column 2 of the same table gives the approximate square-roots of figures given in column 1. Column 3 contains numbers which are obtained by dividing the square-roots by 50. These numbers give the sides of the squares. They can be taken to be so many inches, cms. or any other convenient measure of length. Inches are generally preferred. The four squares in Figure 8 are made to rest on the same horizontal line with equal intervening spaces between them. The squares show the proportionate differences, though it needs a practised eye to detect them.

Table 39 *Production of Cane sugar for* 1938-39

Countries	Quintals 000 omitted 1	Square roots 2	Sides of the squares in inches 3
1 India	[illegible]	[illegible]	1 05
2 Neth Ind Java	[illegible]	39 37	0 79
3 Hawaii	[illegible]	28 90	0 58
4 Columbia	51	7 14	0 14

It will be seen that scale is given above the squares in Figures 8 Let us see how this scale is calculated The side of the first square relating to India is 1 05 inches The area of the square is 1 1025 square inches It represents 2,75,00,000 quintals Therefore 1 sq inch represents 249,40,000 quintals nearly This is the scale to which the squares have been drawn in Figure 8.

Cane sugar production in certain countries 1938-39

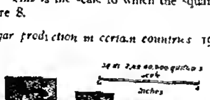


Fig 8

In Figure 8 squares are shown separately one from the other Separate squares do not afford a proper view of proportions at a glance and also require much space If a total is

capable of being divided into parts, the total may be shown by a square and its components sub-divided in it as they are sub-divided in rectangles. This method would not require much space, and would, at the same time, render comparison of proportions easy. For instance, given the area of the world and of the several continents, the area of the world may be represented by a square and the areas of the several continents by its rectangular sub-divisions. The square may be sub-divided horizontally or vertically.

Squares require much time and labour to be drawn accurately. They may therefore be replaced by circular diagrams. Circles take less time to be drawn, and can be drawn sufficiently accurately.

Circular Diagrams—Circles.

The area of a circle varies as the square of its radius. If the radius of a circle is twice that of another, its area would be four times the area of the other. Similar is the case with squares. If the side of a square is twice that of another, its area would be four times the area of the other. It follows that if the radii of several circles are in the same proportion as the sides of the squares, the areas of the circles would also be in the same proportion as the areas of the squares. Therefore, the same quantities may be used either as the radii of several circles or as the sides of several squares, without resulting in any difference in their comparative study. Thus the squares shown in Figure 8 may be replaced by an equal number of circles whose radii would be the same quantities as represent the sides of the squares. Besides, circles are more attractive to look at and, in many cases, more effective to compare than squares. The ease with which they can be drawn is another advantage of circles. Therefore.

when there is a choice between squares and circles, the latter are generally preferred

Table 40 gives the production of ginned cotton in five countries, for 1937-38, in column (2) The square-roots

Table 40 *Production of Ginned Cotton, 1937 38*

Country (1)	Production in Quintals 0000 s omitted (2)	Square roots (3)	Radii in inches (4)
India	1048	32 3	1 53
China	700	26.4	1 27
Brazil	460	21 5	1 02
Peru	81 5	9 0	.43
Argentine	51.4	7 2	34

of figures in column (2) are given in column (3) These square-roots may be taken as the radii of the circles, but since they are too big to be shown in inches or cms. they are divided by 21 The numbers thus obtained are shown in column (4) They stand for radii in inches, and have been used in drawing the circles in Figure 9 The circles are arranged in descending order The proportions of production of ginned cotton in different countries are comparable at a glance In interpreting the circles in Figure 9 we *should not* say that production of ginned cotton is the least in Argentine in the world, since there may be several other countries, not shown in the diagram, which may have still less production

All that can be said is that among the five given countries production in Argentine is the least.

Production of Ginned Cotton in certain countries.

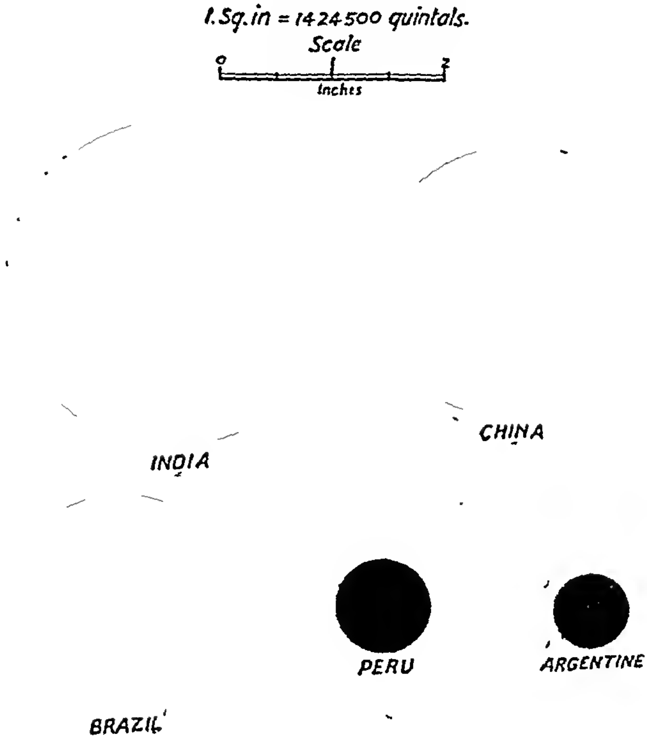


Fig. 9

Angular Diagrams—Sectors.

Just as a rectangle or a square can be sub-divided into its components, a circle can also be divided into sectors to represent the parts of a total. The method of circles and sectors is to be preferred to that of sub-dividing a square, because of the suggestive and attractive character of sectors and circles, and also because of the ease with which they can be drawn.

Table 41 gives the area under food crops cultivated in 1930-40 in the United Provinces and the Punjab

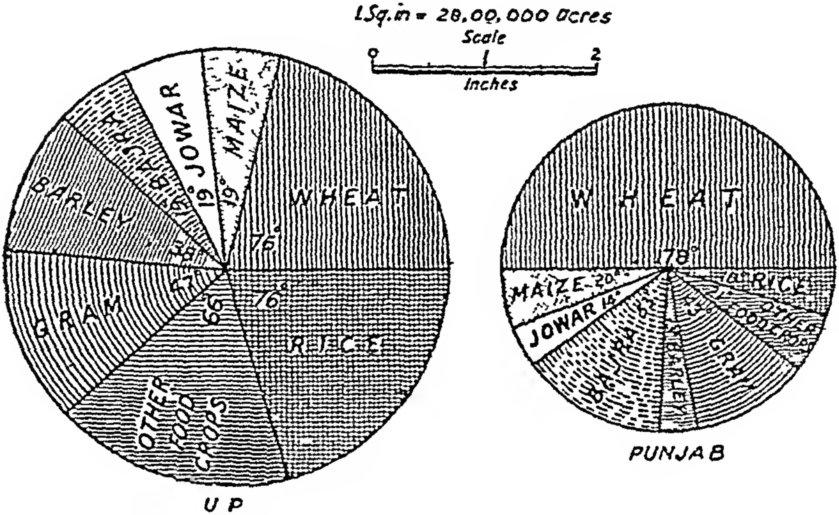
Table 41 *Area under food crops in the U P and the Punjab* 1939 40

Food crop	U P: Area in acres (000 s omitted)	U P: Area in acres (000 000 s omitted)	U P: Angles of the sectors	Punjab: Area in acres (000 s omitted)	Punjab: Area in acres (000 000 s omitted)	Punjab: Angles of the sectors
Rice	[illegible]64	8	[illegible]6°	[illegible]	1	60°
Wheat	810[illegible]	8	6°	95[illegible]	10	1[illegible]°
Barley	[illegible]	4	38°	730	0.7	13°
Jowar	[illegible]307	[illegible]	1[illegible]°	7[illegible]8	0.8	14°
Bajra	2[illegible]33	[illegible]	19°	30[illegible]1	3	53°
Maize	210	[illegible]	19°	1443	1.1	20°
Gram	[illegible]	5	47°	2113	2.4	43°
Other food grains & pulses	6819	7	66°	1931	1.2	1°
Total		38	360°		20.2	360°

Figure 10 represents the data given in table 41 Two circles are drawn one for area under total food-crops in the U P and the other for area under the same crops in the Punjab such that the areas of the two circles are in proportion to the areas under food crops in the two provinces Their radii are in the ratio of $\sqrt{38}$ $\sqrt{20.2}$ To divide these circles into sectors the principle to note is that the areas of the sectors should be proportional to the areas under different crops Now the areas of the sectors are proportional to the angles at the centre Therefore 360° the total number of degrees contained in a circle are to be divided into proportional parts to get sectors of the required areas 38 and 20.2 in the example under consideration are put down as equal to 360° and by the simple rule of three the number of degrees

contained by the angle of the sector representing the area under any crop is calculated. Thus, the angle of the sector for

Area under different food-crops in the U P. and the Punjab



Note.—Figures inside the sectors stand for degrees

Fig. 10

area under rice in the U. P. is equal to $\frac{8}{38} \times 360 = 76°$ nearly. Similarly, angles of other sectors standing for other crops in the two provinces are found out. Having determined the angles of different sectors, we show them at the centre of circle The process would be to draw any radius in a circle, and taking it to be the starting point, to mark off the requisite number of degrees on the circumference with the help of the protractor, and then to draw radii from these marks to divide the circle into requisite number of sectors.

It should be noted that we have proceeded with approximated figures in calculating the angles of the sectors and not with actual figures. The difference between angles found by using the correct and the approximate figures would

be insignificant and immaterial for all practical purposes If greater accuracy is desired the direct method where actual figures are used should be adopted

Scale is given at the top in Figure 10 Its calculation does not involve any difficulty The radii of the two circles are in the proportion of $\sqrt{38}$ $\sqrt{20\ 2}$ that is 6 1 4 5 We have taken 2 03 inches as the radius for the circle for U P, and 1 5 inches for that for the Punjab The area of a circle is equal to $\frac{22}{7}$ therefore the area of the circle for the Punjab is equal to $\frac{22}{7} \times (1\ 5)^2 = \frac{50\ 0}{7}$ square inches It represents 2 02 00,000 acres Therefore 1 square inch represents 28 00 000 acres This is the scale to which circles have been drawn In Figure 10 comparison of the areas under all food crops in the U P and the Punjab can be made at a glance by looking at the two circles and comparison of the areas under different crops can be made by a glance at the sectors and their angles

If the components of a total are too many too many sectors will be required to represent them The circle will in such a case become complicated and lose its effectiveness In order to avoid it figures below a certain quantity may be grouped together so that the number of sectors may be reduced to manageable limit

In Figure 10 it will be seen that in the circle representing the area under food-crops in the U P the sectors are arranged in order of magnitude This arrangement has the advantage that even if minute differences in areas for different crops are ignored in the calculation of the angles of the sectors they will be clearly indicated by the order in which the sectors occur in the circle For instance the areas under rice and wheat in the U P are respectively, 77 64,000 and 81 09 000 acres The difference is very little and has to be

ignored in the process of calculating the degrees of sectors; but this difference is accounted for in the circle by placing wheat before rice. Similarly, the differences between the figures for *jower, bajra* and maize are so slight, that they give equal degrees for their sectors, but by placing *bajra* before *jowar*, and *jowar* before maize the fact that although their degrees are the same their magnitudes are in the order in which they have been arranged in the diagram is made clear. Again, it will be noted that the arrangement adopted in the second circle, namely, that for the Punjab, is just the same as that in the first circle, the reason being that only in this manner would proper comparison between areas in the two provinces under the same crop be made conveniently. It follows, therefore, that when only one circle is drawn to represent components of one magnitude, the sectors in it should *be arranged in some order—ascending* or *descending—and, when* two or more than two circles are drawn, the sectors should have the same order in all the circles.

It is, no doubt, theoretically possible to use circles and sectors for showing the distribution of different incomes over different items of expenditure; but in practice it is not done for two reasons. Firstly, it is easier to draw sub-divided rectangles than to draw sub-divided circles, for the calculation of angles of sectors involves considerable labour. Secondly, the heights of sub-divisions of a rectangle are measurable on the scale showing percentages, and are, therefore, comparable directly in percentages; while the sectors of a circle are comparable directly in terms of angles and only indirectly in terms of percentages. Therefore, rectangular diagram is to be preferred for presenting family budgets to circular diagram. Circle divided into sectors, however, is a good diagram for showing the distribution of world population into various continents or of world area into areas for different continents.

Three Dimensional Diagrams—Cubes

When quantities which have the ratio of 1 1000 are to be diagrammatically represented, even squares and circles to say nothing of bars fail to serve the purpose for if circles or squares are drawn in such a case their radii or sides will have to be in ratio of 1 32 which dimensions are difficult to show on the same scale In cases like this surface diagrams such as squares and circles are abandoned in favour of volume diagrams or three dimensional diagrams, such as cones blocks spheres cubes etc Of course cubes are the easiest to draw and are generally used. The sides of cubes are equal to the cube root of the data to be presented The sides of two cubes representing 1 and 1000 will be in the proportion of $\sqrt[3]{1}$ $\sqrt[3]{1000}$ or 1 10 Cubes should be used only in those cases in which the data cannot be adequately presented through bar or surface diagrams

Table 42 gives the production of tea in certain provinces of India and Coorg The quantities giving the

Table 42 *Production of Tea in some Provinces of India and Coorg in 1939*

Provinces	Production in 000 lbs	Production (Reduced figures)	Cube roots	Side of cube in inches
Bengal	1 12,290	864	9 6	1 9
Madras	38 872	299	6 7	1 3
Punjab	2 807	22	2.8	6
Bihar	1,335	10	2 2	.4
Coorg	130	1	1 0	.2

production are very great, and therefore they are reduced to smaller ones by dividing each of them by 130, the lowest quantity in the table. Cube-roots of these reduced figures are then found out, but since the values of the cube-roots are sufficiently large, each of them is divided by 5 and the resulting figures, given in the last column of the table, constitute the sides of the different cubes.

The construction of a cube requires explanation. Suppose the side of a cube is 1.9 inches, as it is in the case of Bengal. First, a square with its side equal to 1.9 inches is drawn. Then another square of the same size is placeed behind it in such a way that a quarter of each of the two squares just covers each other. This being done, the corresponding corners of the two squares are joined and the construction of the cube is complete. In Figure 11, the method of constructing the cube is given separately. Figures I and II show the position of two squares ABCD and abcd having equal lengths

Production of Tea in 1939 in some Provinces of India and Coorg

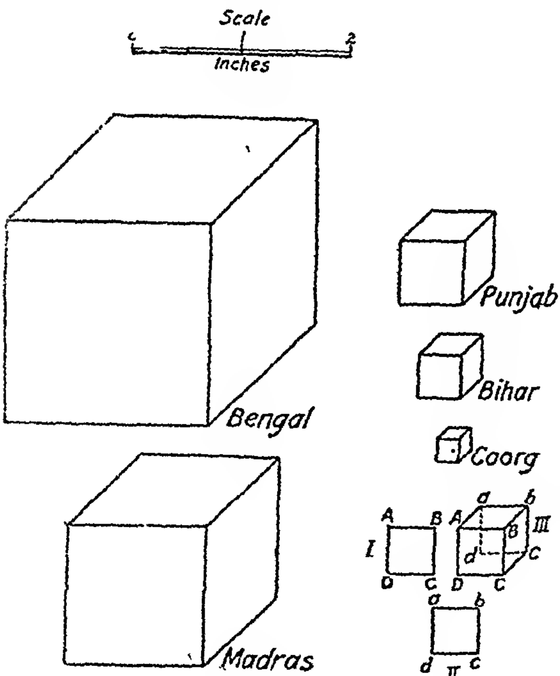


Fig. 11

of their sides abcd is placed over ABCD in Figure III in such a manner that ad and dc are bisected by AB and BC ad and dc are then rubbed off, and the corresponding points A and a, B and b and C and c are joined The sides to be rubbed off are shown by dotted lines in figure III ADCcba is the required cube

In this manner all cubes have been drawn in Figure 11 to represent the data given in table 42 They are all drawn on the same scale and are therefore comparable with one another

Pictograms—Map and Pictures.

The device of pictures is—being profusely used now for comparing statistical data. One comes across pictorial representation of facts very often in co-operative courts of various exhibitions held in the country They are also being

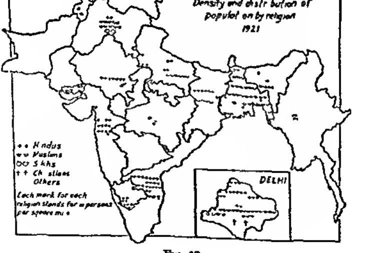


Fig 12

very much used for effective propaganda purposes. The reasons for their coming into popularity are not far to seek. They present dull masses of figures in interesting and attractive manner through objects of daily observation. The image of the entire data is fixed in the mind of the observer by a mere glance at the picture. Relationship between figures and their comparison can be studied through pictograms much more easily than by studying huge mass of numerical data.

Figure 12 shows the density and distribution of population in India by religion. It shows two facts at one and the same time, *viz.*, it shows the density of population in different provinces, and it exhibits the proportions of people belonging to the prominent religions of the country. From a study of the map it will be found that the total number of marks for different religions for the Punjab is 18 which means that the density of population in the province is 180 persons per square mile. Out of these 18 marks, 6 stand for Hindus, 10 for Muslims and 2 for Sikhs. Hence out of 180 persons per square mile in the Punjab 60 are Hindus, 100 are Muslims and 20 are Sikhs. Christians and others being insignificant. It will be seen in the map that the mark chosen for each religion in itself corresponds with the sign of the religion concerned, *e.g.*, the Swastika for the Hindus and the Moon and the Star for the Muslims. This enables us to compare at a glance the proportions of different religions along with the density of population. Instead of using religious signs we could have put down typical human figures representing people belonging to different religions. Of course, the size of the map would have been larger, but at the same time the figure would have been more interesting and attractive.

Figure 13 contains pictures of two money bags, one showing the expenditure on industries and the other that on police by the 11 provincial governments in India in 1940-41.

The expenditures are respectively Rs 115 lakhs and 1120 lakhs. The money bags are not drawn at random. They are enclosed in squares whose areas are in the proportion of 115 : 1120, or whose sides are in the proportion of $\sqrt{115}$: $\sqrt{1120}$, or 10.72 : 33.47 or 1 : 3 approximately. A mere glance at the bags enables a comparison of the amounts spent on the heads of industries and police by provincial governments in India.

The money bags could have been placed inside bars or rectangles whose width may have been equal and heights in proportion to the amounts. But then the shapes of the bags would have looked unnatural and instead of being attractive the pictures would have become repulsive. Great caution is, therefore, necessary in presenting statistical data in pictorial form. Pictures cannot be used indiscriminately. Serious thought must be devoted before using them, lest they might look unnatural or ridiculous or mis-represent the data. Pictures are usually enclosed in squares, circles or rectangles for the reason that it is not possible to calculate the surface area of pictures. The pictures of the bags for example are very irregular so that their areas cannot be easily determined. They have therefore been put into squares.

Pictogram showing expenditure of the eleven provincial governments in India in 1940-41 on Industries and Police

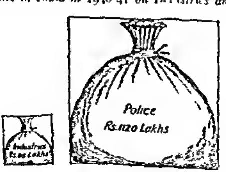


Fig 12

General Remarks.

Several devices of representing statistical data diagrammatically have been considered above. Forethought with regard to the suitability of a particular form in a given case and practice in drawing diagrams are essential factors. Bars and circles, it will be found by experience, are the easiest to draw and suitable for general use. A particular case, however, may necessitate a square, a rectangle or a cube. Attention, in drawing diagrams, should always be fixed upon their neatness and on the precision with which they represent facts. As has been already said, continuous changes over a period of time are best shown by graphs, and not by diagrams. Diagrams should therefore be used in representing spatial series.

EXERCISES

(1) What is meant by diagrammatic representation of facts? What is its importance?

(2) How far is diagrammatic representation an advantage over statistical tables?

(3) What are the different forms of diagrams? Explain in detail the construction of any two of them.

(4) What precautions are necessary in drawing a good diagram? What is the test of a good diagram?

(5) What mistakes are generally found in diagrams? How would you avoid them?

(6) Show with the help of a few examples that diagrams can be wrongly used.

(7) Write short notes on:

Bar-diagram, pie-diagram, three-diamensional diagram, pictogram.

(8) What kind of statistical data are best represented by diagrams? Illustrate your answer with examples.

(B. Com., Agra, 1937).

(9) Illustrate the following by suitable diagrams —

(a) In 1931 twelve seers of wheat could be had for one rupee while in 1943 a rupee purchased only two-and a half seers of wheat.

(b) 120 out of every 1,000 of the population of India were literate in 1941 as against 95 ten years past

(c)	I	II
Price per unit of a commodity	Rs 10	Rs 12
Quantity sold	20	24
Value of raw materials used	Rs 100	Rs 120
Other expenses of production	60	, 96
Profit	40	„ 72

(10) Diagrams are meant for a rapid view of the relation of different data and their comparison —Discuss

Draw a 'Bar or Pie' diagram to represent the following data —

Output and Cost of Production of Coal

	1924	1928
Cost per ton disposed commercially		
Wages	12 74	7.95
Other Costs	5.46	4 51
Royalties ..	0.56	0.50
Total	18.76	12.96
Proceeds of scale per ton	19 91	12 16
Profit (+) or loss (—) per ton	1 15	—0.80

(M A., Agra 1940)

(11) Represent the following data by suitable diagrams:—
Electricity sold in India in 1940-41 and 1941-42

	1940-41 units (000)	1941-42 units (000)
Domestic consumption ..	156,916	138,301
For offices and other uses	103,503	109,805
Industrial Power ..	1,373,631	1,603,487
Street lighting ..	46,419	32,563
Tramways ..	37,471	46,315
Electric Railways ..	161,534	315,223
Miscellaneous ..	60,880	110,934
Total units sold ..	1,940,624	2,356,628

(12) The following figures relate to the postal traffic in India in 1940-41 and 1941-42. Represent them by suitable diagrams.

Articles	1940-41	1941-42
	Number (000)	Number (000)
Letters	529,096	541,528
Postcards	365,458	413,096
Regd. Newspapers ..	78,535	80,578
Books and pattern packets	110,703	99,613
Unregistered Parcels ..	3,324	3,426

(13) The following is the distribution of scholars in institutions for females in India in 1939-40.

Assam	..	..	54,891
Bengal	..	..	519,735
Bihar	..	..	82,891
Bombay	..	..	248,880
C. P.	..	..	56,549
Madras	..	..	452,059
N. W. F. P.	..	..	18,977
Orissa	..	..	19,878
Punjab	..	..	202,180
Sind	..	..	40,368
U. P.	..	..	159,099
	Total	..	1,894,590

Represent the above data by suitable diagram

(14) The following are the percentages of expenditure on education in 1939 40 for three provinces

Sources	Assam	Bengal	Bihar
Govt Funds	54 64	34 20	29.47
Local Funds	13 09	7 30	28 39
Fees	21 66	42 70	28 04
Other Sources	10 61	15 80	14 10

Represent the above figures by suitable diagram

(15) Value of the imports of glass and glassware into India from different countries—during the year 1931 32

Japan	42 lakhs of rupees
Czechoslovakia	23
Germany	20
U K	13
Belgium	13
Other countries	11

Represent the above figures by suitable diagrams

(B Com Alld 1933)

(16) Draw a simple diagram to represent the following statistics relating to the area under different crops in India in 1933 34 and write a brief note on the given data —

Crop	Million Acres
Rice	80 3
Wheat	27 6
Jowar	21.4
Other food crops	88 2
Oil seeds	17 8
Cottons	14 5
Other fibres	3 1
Fodder crops	10 2
Other non food crops	39

(B Com Cal 1937)

(17) The following table gives the birth rates and death rates of a few countries of the world during the year 1931:—

Country	Birth rate	Death rate
Egypt	44	27
Canada	24	11
U. S. A.	19	12
India	33	24
Japan	32	19
Germany	16	11
France	18	16
Irish Free State	20	14
United Kingdom	16	12
Soviet Russia	40	18
Australia	20	9
New Zealand	18	8
Palestine	53	23
Sweden	15	12
Norway	17	11

Represent the above figures by a suitable diagram.

(B. Com., Luck., 1938).

(18) The following table gives the details of monthly expenditure of three families:—

Items of Expenditure	Family A Rs.	Family A As.	Family B Rs.	Family B As.	Family C Rs.
Food	12	0	30	0	90
Clothing	2	0	7	0	35
House-rent	2	0	8	0	40
Education	1	8	3	0	12
Litigation	1	0	5	0	40
Conventional necessity	0	8	3	0	60
Miscellaneous	1	0	4	0	23

Represent the above figures by a suitable diagram. Which family is spending the money most wisely? Give reasons.

(M.A., Alld., 1937).

(19) The following table gives the details of the cost of construction of a house in Allahabad —

	Rs
Land	4 500
Labour	2 500
Bricks	2 000
Iron	1,800
Timber	1,500
Cement	800
Lime	800
Stone	600
Sand	200
Other things	1 300

Represent the above figures by a suitable diagram

(B Com Alld 1945)

(20) Represent the following by subdivided bars drawn on percentage basis —

	1933 Rs	1938 Rs	1945 Rs
Proceeds per chair	12	12	20
Costs per chair			
Wages	7	7	10
Other costs	5 5	4	6
Polishing	1 0	1 5	2 5
Total	13 5	12 5	18 5
Profit or loss per chair	—1 5	— 5	+1 5

(21) Illustrate the following by suitable diagram —

Production of Cotton in Egypt

1911—12	3111000 tons
1914—15	6450000 tons
1918—19	10873000 tons

(22) Following are the figures of the population of the various countries of the world and of total world population in 1931 —

Country	Population (000 s omitted)
China	411 770
India	352 370
U S S R	161 000
U S A	124 070
Germany	64 776

Country	Population (000's omitted)
Japan	64,700
U. K.	46,077
France	41,860
Italy	41,100
Others	705,077
World	2,012,800

Represent the above figures by a suitable diagram.

(23) Illustrate the following data diagrammatically:—

Area in (000) square Kilometers of the continents of the world.

Continent	Area
Asia	41,900
S. America	40,687
Africa	29,946
N. America	19,653
Europe	11,426
Oceania	8,550
Others	2,764
World	154,926

(24) Represent the following figures diagrammatically:—

City	Females per 1,000 males
Calcutta	464
Madras	908
Lahore	596
Lucknow	516
Banaras	781
Peshawar	708
Tinnevelly	1,068

(25) Show diagrammatically the balance of trade in cotton piece-goods from the following data.

	Exports million yds.	Imports million yds.
1939-40	221.3	579.1
1940-41	390.1	447.0
1941-42	779.4	181.5

Also show in separate diagrams

(a) the proportion of exports and imports to total trade in 1941 42

(b) the proportion of exports to imports in 1939-40

(26) Represent the following by a suitable diagram

Principal Heads of Revenue	1938-39 Lacs of Rs	1939 40 Lacs of Rs
Custom	4050	4588
Central Excise Duties	868	652
Corporation Tax	204	238
Taxes on Income	1374	1420
Salt	812	1080
Opium	50	46
Other Heads	112	130

(B Com Nagpur 1943)

(27) Draw suitable diagrams to represent the following information —

(a) Town	Literate	Illiterate	Total Population
A	1,000	9 000	10 000
B	160 000	200 000	360 000

(b) Factory	Wages Rs	Materials Rs	Profits Rs	Units Produced
A	2 000	3 000	1 000	1 000
B	1,400	2,400	1 000	800

Show also cost and profit per unit

(B Com Agra 1943)

CHAPTER XV

GRAPHICAL PRESENTATION

Diagrams and maps discussed in the preceding chapter are particularly suited to the comparison of variables spread over different places or different heads at the same time, that is, they are useful for representing spatial series. For illustrating series spread over a period of time and also for illustrating frequency distributions, graphical methods are made use of. In this chapter, therefore, we shall discuss

(1) Graphs of Time Series showing continuous changes, and

(2) Graphs of Frequency Distributions.

Diagrammatic and Graphic Presentations Contrasted.

In the diagrammatic method bars, rectangles, circles etc. stand for quantities individually or in groups. In the graphical method quantities are not *represented* by one or more dimensional figures, but are *located* on a surface with respect to two or more dimensions, for which purpose a system of rectangular co-ordinates, such as that given on page 336, is used.

Two straight lines x x' and y y' intersecting each other at right angles are drawn. The horizontal line is the x-axis or *abscissa* and the vertical line the y-axis or *ordinate* The junction of the axes at o is known as the point of origin or zero. Distances measured towards the right or upwards from the origin are reckoned as positive, and those measured towards the left or downwards as negative. All points in the plane, *i.e.*, the four quadrants into which the graph is divided, are located by reference to two co-ordinates drawn parallel to the axes.

The scales of measurement on the axes may be chosen at convenience There is no necessary connection between

A System of Rectangular Co ordinates

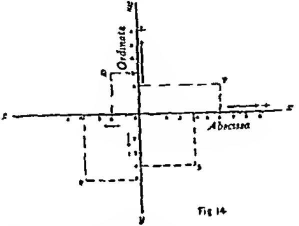


Fig 14

the x and the y scales In Figure 14 the scales of x and y axis are equal and the points P Q R S have been located as follows —

Point	x	y
P	+6	+2
Q	—2	+3
R	—4	—5
S	+4	—4

The co-ordinates on the points are indicated by dotted lines But the dotted lines need not be made visible, only the point should be shown In practice graph papers which have a net work of fine lines, are used for graphical presentation. By using them the necessity of drawing lines is dispensed with

It is evident from the above that to *locate* quantities in a plane bounded by the x and the y axes in the manner illustrat

ed in Figure 14 is not the same thing as to *represent* them by bars, squares, rectangles, etc. The latter are themselves drawn proportional to the amounts which they represent.

All truly continuous series are properly presented by graphical as distinct from diagrammatic methods. Graphic methods are more powerful than diagrammatic ones in so far as they not only present the facts effectively but also bring to light new relations that may not at first sight be visible from a study of the quantities themselves, for instance, we may determine through graphs whether phenomena are connected or independent.

GRAPHS OF CONTINUOUS TIME SERIES

Rules for drawing Graphs.

A continuous series may be measured (1) in time, or (2) in space, or (3) be represented by frequencies of a variable at the same time or place. We shall first deal with the graphical presentation of the first type of series, *viz.*, time series.

Time series is also called **historical series,** since it stands for the numerical record of the changes in a variable during a number of successive intervals in a period of time. The first problem to be considered in drawing a graph of such series is the choice and adjustment of scales.

Choice and adjustment of scales.—A system of rectangular co-ordinates as illustrated in Figure 14 is used to illustrate time series. Time units are placed on the abscissa or x-axis, and the sizes of variable measured on the ordinate or y-axis. Since time has no zero, the horizontal scale need not begin with zero; the first time interval may be indicated near or away from the point of origin or o. The x-axis should be divided into equal parts, each of which should represent periods of equal length. The vertical scale should begin with zero when sizes of variable are shown on it, since they are

always reckoned from zero Equal space units on the y axis should represent equal amounts when natural scale is used and equal rates of change when ratio scale is used Just now, we shall be concerned with the natural scale

What should be the proportion between the abscissa and the ordinate scales? Bowley states the problem and the way in which it should be solved as follows —

> It is difficult to lay down rules for the proper choice of the scales by which the figure should be plotted out It is only the ratio between the horizontal and vertical scales that need be considered The figure must be sufficiently small for the whole of it to be visible at once if the figure is complicated relating to a long series of years and varying numbers minute accuracy must be sacrificed to this consideration Supposing the horizontal scale decided the vertical scale must be chosen so that the part of the line which shows the greatest rate of increase is well inclined to the vertical which can be managed by making the scale sufficiently small and on the other hand, all important fluctuations must be clearly visible for which the scale may need to be increased Any scale which satisfies both these conditions will fulfil its purpose [1]

Thus the scales chosen should be such as would allow the full data to be presented on the graph would properly show the extreme fluctuations and would clearly bring out the changes over the entire period from date to date The two scales selected will no doubt depend upon the size of the paper still it should be carefully noted that if the units occupy too much space small changes in the size of variable will appear to be important fluctuations while if the units occupy too little space even large fluctuations would look unimportant. The respective scales will, it is obvious be different for different data No one standard can suit all cases, yet, it is

[1] Bowley A L, *Elements of Statistics* 1920 ed p 132.

desirable as a general rule, to have the x-axis approximately $1\frac{1}{2}$ times as long as the y-axis.

Having decided the proportion of the two scales, the ordinate scales should be divided into units such that they would be easily comprehended in terms of the rulings of the paper used. For instance, if the paper is ruled in fifths or tenths, ten small squares should not be made equal to such an amount as 4357. A given space should equal some multiple of ten or five, as 3000, 250, 25 etc. When the scale has been divided into units, the ordinate should be labelled in terms of the scale unit, *i.e.*, at each equal space the value of it in terms of the scale unit should be put down. One should not try to fill up the scale with too many details, and the putting down of each successive frequency—or every frequency that is plotted—should be avoided.

The natural scale is thus ready. We shall deal later with ratio scales and false base line. Presently, the problem of plotting the data is taken up.

Plotting the data.—To plot the given data the method shown in Figure 14 will be followed. To plot the size of an item against a particular date, only a point placed on the ordinate concerned is enough. Thus when the whole data are plotted, there will be as many points on the plane as the number of dates. Since time is a continuous factor, these several points should be connected from date to date by continued smoothed lines, each point being simply a conventional stopping place. This continuous smoothed line is called the **curve;** it shows the probable changes at all possible intervals of the entire period to which the data relate. This curve is also given the name of **Historigram.** This name must be distinguished from histogram into the construction of which the factor of time does not enter. But drawing a smooth curve requires practice and skill which everyone does not usually possess.

An alternative which is commonly adopted therefore is that of connecting the points plotted by *straight lines** When the curve is plotted it should be given a short but adequate title indicating what it represents And if several curves are plotted each should be differentiated either by using different inks or by adopting some such devices as drawing straight continuous line dotted line dot and bar line if using the same ink An explanation of what the different inks used or lines drawn indicate in the historigram should be separately given in a corner of the paper as given in Figure 16 page 344 or the name of the factor represented may be written on the curve itself if it does not spoil the figures as done in Figure 30

Different Types of Graphs on the Natural Scale

The purpose of graphical presentation as that of diagrammatic representation is comparison Comparison is necessary to have a clear idea of the relationship of things in time and space The methods by which comparison in space is rendered possible have already been considered in the preceding chapte Here the purpose is to consider the methods by which comparison in time is made Our aim may be to study

1 Changes of a single variable over a period of time and

2 Changes of two or more variables over a period of time

These changes can be studied on the Natural Scale or Difference Charts and on the Ratio Scale or Ratio' Charts

First we take up the natural scale graphs On it changes of a single variable are studied through (1) Absolute historigram and (2) Index historigram And changes in two or more variables are also compared through (1) Absolute historigrams and (2) Index historigrams In addition to these we shall take up a discussion of a few other ways of drawing graphs for comparison of different variables and of the false base line

* In Mathematics the term curve includes a *straight* line

Absolute Historigram of one variable.—When original quantities, and not index numbers, are presented graphically, the resulting figure is called 'absolute historigram' as distinguished from an 'index historigram' which relates to an index number series. In Figure 15, amounts of treasury bills tendered in India from week ending 10th February, 1942, to week ending 28th April, 1942, given in table 43, are shown. One small division represents on the x-axis one week, and on the y-axis Rs. 5,000,000. These scales give an outline which is neither too flat nor too angular and shows the fluctuations clearly. Since all the quantities in table 43 are positive, only the north-east quadrant of the system of rectangular co-ordinates has been utilized. Only one variable has been treated in the figure and changes in its size are comparable through the graph from week to week. Thus, the amount of treasury bills tendered increased from Rs. 30,000,000 on 17th March, 1942, to Rs. 33,000,000 on 24th March, 1942 that is, it increased by Rs. 3,000,000.

Table 43. *Treasury Bill Tenders in India.*

Week ending 1942 x		Amount tendered Rs. (000,000) y
Feb.	10th	22
"	17th	28
"	24th	27
March	4th	23
"	10th	24
"	17th	30
"	24th	33
"	31st	31
Apr.	7th	30
"	14th	24
"	21st	27
"	28th	35

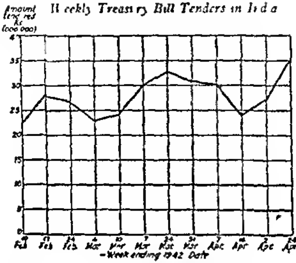


Fig 15

Index Historigram of one variable.—When a time series consisting of index numbers is given it may be graphically presented in just the same manner as the series consisting of absolute or actual values In Figure 28 an index historigram relating to the retail price of wheat in India is shown along with other curves An *important difference* between index historigram and absolute historigram is that the latter shows the actual values and therefore studies the *movement* of absolute sizes of the variable from date to date while the former shows the index numbers and therefore studies the percentage change on a particular date as compared with the *base* year Thus from Figure 28 we can study that retail price of wheat in 1877 was two per cent more than in 1873 the base year we cannot study the absolute amount of money by which the increment was effected

Absolute Historigrams of two or more variables (Homogeneous units).—If *two or* more variables measured in

the same unit are given, all of them can be exhibited on the same graph, with *common* vertical and horizontal scales. Figure 16 on page 344 shows values of monthly exports, imports and balance of trade of India for 1932-33. The three curves are drawn in the same manner as shown in Figure 14. It should be noted that since some of the values in the balance of trade series, shown in column 4, table 44, are negative, the eastern or the right half of the system of rectangular co-ordinates has been utilized.

Table 44. *Imports, exports and balance of trade of India during* 1932-33. (*In crores of Rs.*)

Month x	Imports y	Export y	Balance of trade y
April ..	13	11	—2
May ..	12	10	—2
June ..	12	10	—2
July ..	11	9	—2
August ..	11	10	—1
September ..	11	13	+2
October ..	10	12	+2
November ..	11	12	+1
December ..	10	13	+3
January ..	11	12	+1
February ..	9	12	+3
March ..	11	13	+2

In Figure 16 (see next page) comparison can be made only of the absolute amounts of exports, imports and balance of trade, since the historigrams show absolute values. It will not be possible to compare proportional changes between them during the same period.

Absolute Historigrams of two or more variables, (Heterogeneous units).—If two or more variables are measured

in different units all of them can be exhibited on the same graph, but with *different* ordinate scales the horizontal scale being common The ordinate scales will have to be different for the simple reason that the variables are not measured in the same unit With this difference only the curves shall be prepared in just the same manner as they have been done in

Values of Monthly Exports Imports and Balance of Trade of India in 1932 33

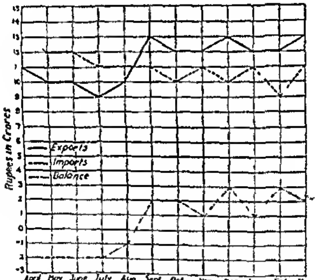


Fig 16

Figure 16 A study of the changes in the absolute amounts of the same variable from one date to another will be possible, but comparison of changes in one variable with those in the other during the same period will not be possible For of two variables if one is measured in tons while the other in yards, tons will be comparable with tons and yards with yards but not tons with yards.

Index Historigrams to compare changes of two or more variables.—In the case of absolute historigrams of two or more variables expressed in the same or different units, it has been noticed that only absolute sizes are comparable and it is not possible to study from them whether change in the value of one variable from one date to another is similar to or different from the change in another variable during the same period. For instance, in Figure 16, it is possible to note that between April and May exports fell from Rs. 11 crores to Rs. 10 crores and imports from Rs. 13 crores to 12 crores. That is, in both the cases the fall is by an equal amount of a crore of rupees. But, is the proportional decline the same in both the cases? This information cannot be had from the figure referred to. To compare proportional change, an easy way will be to reduce the given variables to index numbers on the *same base*, and then plot index historigrams. From the curves thus drawn it would be easy to compare the proportional changes in relation to the base in the different variables during the same period. If figures relating to exports and imports during May, to refer to table 44, are converted into index numbers, it is found that with April figures equal to 100, the index for exports for May is 91, and that for imports for the same month is 92. Thus, it is possible to see that the proportional decrease from April to May in exports is greater than that in imports. The fact that the absolute decrease in both, exports and imports, is the same in no way obscures the record of the comparative proportional change. In index historigrams it is advisable to draw a line parallel to the x-axis from the point 100 on the ordinate, so that proportional change on any date from the base, whose value is put down at 100, may be seen at a glance.

But, as was noticed in dealing with index historigrams of a single variable, index historigrams are no improvement on the absolute historigrams if it is desired to compare different periods

in the same series with regard to the relative changes therein. All index numbers compare the change with the base year and not between themselves. Therefore index curves of exports and imports with April as the base in the example under consideration shall exhibit the comparison of proportional change with April; study of proportionate changes from August to September or from December to January will not be possible. The function of clearly exhibiting relative changes over periods of time is reserved for logarithmic historigrams which we shall study later.

Methods of scale conversion for comparing changes in two or more variables.—It has been noted above that when variables are expressed in different units or even in the same unit they may be converted into index numbers to compare proportional changes relative to the base. Several methods

Table 45 *Volume and Value of Exports of Lac from India in 1941-42*

Month x	Volume y	Value z
	Cwts. (000)	Rs. (00 000)
April	53	72
May	80	34
June	89	40
July	96	50
August	56	35
September	69	43
October	32	23
November	60	48
December	22	19
January	102	83
February	60	61
March	49	46

of comparing the differences between two or more variables during a particular period are, however, available. These relate to converting one scale into terms of the other scales. Of these methods we take up one below, which is particularly suitable in a case where variables are expressed in different units.

The method is very simple. When two or more variables are given, separate scales may be chosen for different variables, but each should be made proportional to the respective averages of each. Table 45 gives the volume in cwts. and value in Rs. of lac exported from India month by month during 1941-42. First, averages of these two series are computed which are, respectively, 64,000 tons and Rs. 41,000. These average values are plotted in Figure 17 on two separate

Volume and Value of Exports of Lac from India in 1941-42.

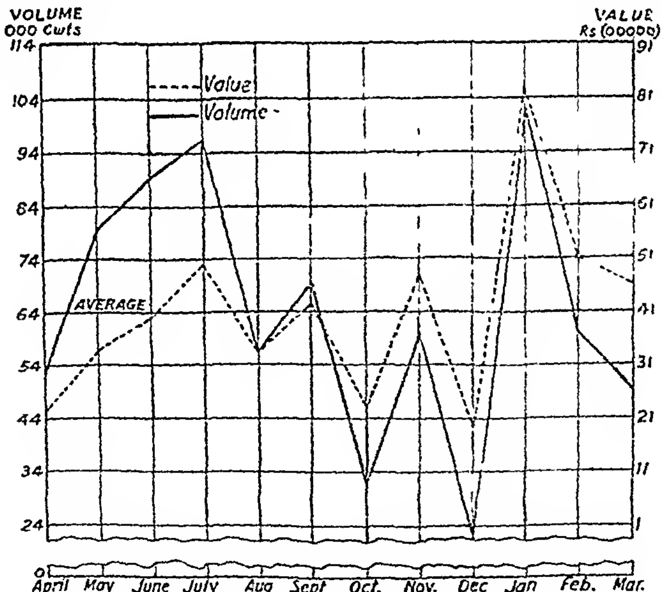


Fig. 17

vertical scales in such a way that the average values of the

two i.e. represented by the same position on the vertical scales. After the scales have been thus adjusted, the values of the variables are plotted. Each of the two curves should be read in terms of its own scale.

One difference between this method and the method of index historigrams discussed above is that in this method actual values are plotted while in the other method values relative to the base year—index numbers—are plotted. In Figure 17 it is easy to compare the two series since their averages lie on the same line.

False Base Line.—In Figures 15 and 16 the scales on the ordinate beginning from zero are continuous while in Figure 17, the scale is broken or discontinuous. In Figure 17 we have used the false base line.

In those cases in which the fluctuations are small relatively to the size of the variable and the insignificance of those fluctuations is to be visualized, and in those cases where adjustment of scales is not otherwise possible (as in Fig 17), instead of showing the entire vertical scale from zero to the highest value involved only as much of it is shown as is just sufficient for the purpose. That portion of the scale which lies between zero and the smallest value of the variable is omitted. This has been done in Figure 17 and is further illustrated in Figure 18 which shows the index numbers of prices of Government Securities during 1937-42. The range of the series is (123—104)=19 and the fluctuations are small. To amplify these fluctuations the vertical scale shown is only between the limits 103 and 124. On this wide vertical scale the fluctuating movement in the price of Government Securities is brought into prominence. The fact that a false base line has been taken and the vertical scale is not shown in its entirety should always be made clear on the graph by the double saw tooth line, as shown in Figures 17 and 18 or by

some other device. If a continuous vertical scale beginning from zero and going up to the maximum value of the index

Index numbers of prices of Government Securities during 1937-42.

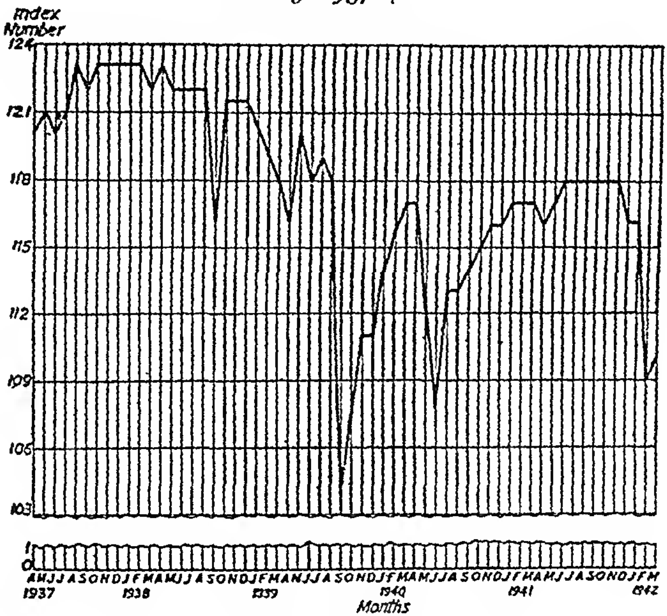


Fig. 18

numbers were used to plot the indices of the prices of Government Securities, a slightly waving line would have resulted and the fluctuations would have been quite inconspicuous. Similarly, in Figure 17 adjustment of the scales would not have been possible without resorting to the false base line.

False base line should be used in rare and exceptional cases. It is always safe to show the zero line, for a correct study of the proportional changes is possible only when the zero of the ordinate in the graph is shown and the height

of every point on the curve is shown in full For exampl the proportion between two numbers 100 and 400 if expressed in full above the zero line is 1 4 But if the vertical scale from 0 to 50 is omitted each of the two figures will be reduced by 50 If the horizontal line passes through 50 the respective magnitudes of these two numbers above the horizontal line will be 50 and 350 whose ratio will not be 1 4 but 1 7 Similarly if the vertical scale is omitted from 0 to 70 the two quantities 100 and 400 will be reduced to 30 and 330 respectively above the base line passing through 70 of the vertical scale Their ratio will no longer be 1 4 but would be changed to 1 11 Thus wrong impressions might be created in regard to proportions if a portion of the vertical scale is omitted i e if a false base line is taken This would however not happen when the significance of the false base line is kept in view

Not infrequently people resort to false base line either to economise space or for want of space Thereby they unconsciously run into the error of making fluctuations look larger than they really are Sometimes this is done deliberately *too* Therefore while studying curves and commenting on them this fact must not be ignored

Graphs on "Ratio" Scale

So far we have been dealing with the natural scale graphs in which the y's are scaled in proportion to their actual values We have seen that this method shows *absolute* movements in statistical series but fails to exhibit relative movements in their proper perspective. The importance of the study of relative changes in economic investigations is grow

ing. To study relative changes the Ratio scale or **Logarithmic scale** is employed as an alternative to the Natural scale.

Natural Scale contrasted with Ratio Scale.

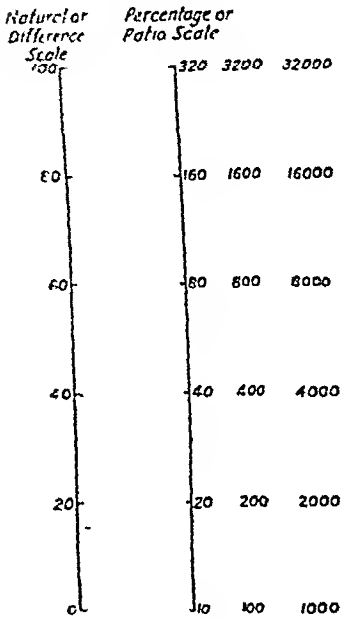


Fig. 19

The difference between these scales is that with the natural scale equal distances on the ordinate represent equal *absolute* movements, while with the ratio scale they represent equal *proportionate* movements.

Ratio Scale.—Ratio scale is based on geometric progression, while the natural scale is based on arithmetic progression. This fact is borne out in Figure 19, which compares the natural scale with the ratio scale.

The importance and usefulness of the ratio scale can be very easily seen. Let us suppose that the population of a certain town increased as follows

Year	Population	Increase
1920	200,000	
1930	300,000	50 per cent.
1938	400,000	33·3 per cent.

The absolute increase in the two periods under study is identical, that is, 100,000 in each case, but the proportional increase differs, for in the first case it is 50% of 1920, and in the latter, $33\frac{1}{3}$% of 1930. If then, these population figures are plotted on the natural scale, the increment in population (100,000) in

each case will be shown by equal distances on the vertical scale thereby leading to the conclusion that the increments between 1920 and 1930 and between 1930 and 1938 were equal. But the ratio graph will indicate that the increase took place at the rate of 50% and 33⅓% respectively

Logarithmic Curves.—Rates of changes may be graphically presented in either of two ways

(1) by plotting the logarithms of the amounts on a natural scale

(2) by plotting the amounts themselves on a logarithmic scale.

In the former method logarithms of the variable to be plotted on y axis of the natural scale graph are found out plotted and joined by straight lines to result in the required logarithmic curve

In the latter method logarithms of the y variable are not found Rather the actual values of the variable are plotted on logarithmic paper which is ruled on the ratio scale principle, illustrated in Figure 19 This method is more simple and easy than the former and has the additional merit of comparing actual values But if logarithmic paper is not available the former method has to be preferred

Logarithmic curve is also termed as *Semi-logarithmic curve* for the simple reason that one variable y, is plotted on the logarithmic scale, while the other variable *x*, remains on the natural scale

Table 46 shows how a sum of Rs 10 borrowed by A and another sum of Rs 100 borrowed by B in 1934 increase at the rate of 20% per annum compound interest being charged on both the sums

Figure 20 shows a graph of the two series on the *natural* scale. From the figure it appears that the rate of increase in series B is more rapid than that in series A This, however, is not the case

Table 46. *Increment of of sums of Rs 10 and Rs. 100 at 20% p. a compound interest.*

Year	A	B
	Rs.	Rs.
1934	10	100
35	12	120
36	14.4	144
37	17.3	172.8
38	20.7	207.4
39	24.9	248.9
40	29.9	298.5
41	35 8	358.2
42	42.2	421.6

Sums of Rs 10 and Rs 100 rising at compound interest on the Natural Scale.'

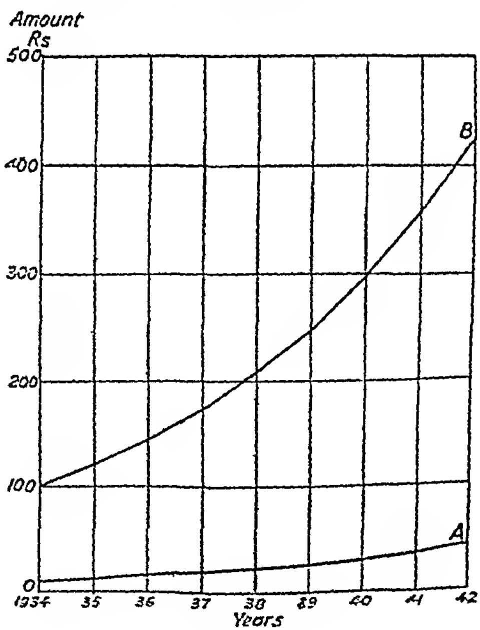


Fig. 20

as is shown in Figure 21 in which the two series are drawn on the *ratio* chart The curves A and B in the figure run parallel to each other and this clearly brings out the fact that the rate of increase in both the series is identical

It will be seen that the curves in Figure 21 have been drawn according to the first of the two methods discussed above *viz*, by plotting the logarithms of the two variables against time on the natural scale graph paper Instead if logarithmic paper had been used and actual items were plotted on it the resulting curves again would have run parallel to each other indicating a uniform rate of increase

Sums of Rs 10 and Rs 100 rising at compound interest on the Ratio Scale

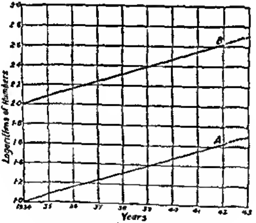


Fig 21

Instructions for reading of Logarithmic Curves—The following general rules will help the reading of logarithmic historigrams —

1 If a curve rises upwards the rate of growth is increasing (*cf* Figure 21)

2. If a curve is falling downwards, the rate is decreasing.

3. If a curve is ascending but is *nearly* straight, the magnitude it represents is growing at a nearly uniform rate.

4. If a curve is descending but is *nearly* straight, the magnitude it represents is decreasing at a nearly uniform rate.

5. If a curve is a straight line the rate of change is uniform or constant. (*cf.*, Figure 21).

6. If a curve is steeper in one portion than in another portion, the rate of change in the former is more rapid than that in the latter.

7. If two curves on the same ratio chart are found running parallel they represent equal percentage rates of change (*cf.*, Figure 21).

8. If one curve is steeper than another on the same ratio chart, the rate of change in the former is more rapid than that in the latter.

When comparison is made between percentage of increase and percentage of decrease directly, it is essential to remember that a loss of 50% is not made good by a gain of 50%, but by that of 100%. Great care should, therefore, be exercised in comparing increases and declines on ratio charts. However, if increases are compared with increases and decreases with decreases no such caution is required.

Advantages and disadvantages of Ratio Scale.—A ratio scale has no zero since it compares relative rates of change. A natural scale has a zero because it compares absolute values. Consequently, zero line is necessary in the natural scale but not so in the logarithmic scale.. Since there is no zero in a ratio scale, there is no danger of fallacious conclusions being drawn from the graph. We have seen on page 350 how fallacious conclusions might be drawn from graphs drawn or natural scale whose zero line is omitted.

Ratio scale makes **extrapolation**—finding out a future probable figure—possible if the data are organic in character. If population figures of a certain country are given and are plotted on the ratio scale, the curve may be extended in continuation with its trend beyond the last date to a next date to obtain thereby a fairly accurate estimate of what the next figure is likely to be. In Figure 21 the curves A and B showing the amount at compound interest have been extended in continuation with their trends and it is possible to read from the dotted line that the amounts at compound interest in 1943 will be Rs 50.6 and Rs 506.2 respectively.

The logarithmic scale is specially important in the case of **index historigrams.** They should generally be drawn on ratio scales because index numbers are more concerned with proportionate changes than with actual ones. Index numbers plotted on the natural scale convey false impression. For example, if price index numbers for three successive years are 100, 130 and 160 each succeeding number differs from the preceding one by 30. This difference would be represented by equal distances on the natural scale so that the rise in prices would appear to be equal. But a change from 100 to 130 implies an increase of 30% while that from 130 to 160 means a rise of 23 1/13 per cent. Therefore the difference between the indices when plotted on the ratio scale would be shown as 30% between the first and the second year and 23 1/13 per cent between the scond and the third year. The percentage changes of the two periods will then be comparable. Such comparison it was pointed out while dealing with index historigrams drawn on natural scale[1], is not possible when simple and not logarithmic index historigrams are used. Thus logarithmic graph is very useful for relative comparisons in point of time.

---

[1] *Cf* page 346

Logarithmic graphs, however, have two disadvantages. Firstly, they are no good for comparison of the absolute sizes of different variables. Secondly, negative values cannot be shown on the logarithmic scale. It may also be added as a third disadvantage that the ordinary reader is unfamiliar with logarithms and logarithmic graphs and is, therefore, unable to interpret what ratio charts really imply.

General Remarks.

We have studied the graphical methods by which continuous series spread over a period of time can be presented and changes in a single variable or in two or more variable can be studied. It should be noted that we have taken no account of the study and comparison of short-time and long-time changes in a time series. It would be taken up in the next chapter.

We have also discussed that the changes in a time series can be graphically studied in their absolute values through absolute historigrams, in their values relative to the base by ndex historigrams, and in their proportionate values by historigrams—absolute or index—drawn on the ratio scale. We have, therefore, studied the methods by which comparisons can be made in point of time. It is now proposed to deal with the graphical methods by which the statistical nature of frequency distributions can be studied.

FREQUENCY GRAPHS

Frequency distribution may be discrete or continuous. A table giving frequency distribution of a group presents the data in compressed form, but many people can normally appreciate the relative sizes of a number of quantities more readily when they are graphically presented than they can do by looking at a table. Graphical presentation may, therefore, be very well

employed as an addition to the method of tabulation in bringing out the statistical nature of a group whether discrete or continuous

Statistical Nature of a Group

In Chapter X a good number of tables relating to frequency distribution of groups are found In all of them one characteristic would be observed It is that frequencies rise up to a certain maximum point and begin to fall after this point is reached Another point to be noticed is that this rise and fall shows certain regularity The question that naturally arise is whether this regular rise and fall noted in the tables in Chapter X is simply an arbitrary assumption a characteristic of the particular frequency distributions referred to or a feature common to many varieties of phenomena It would be found that this feature is common to many or most phenomena

Let us pick up a good number of leaves from any tree *at random* measure their lengths and tabulate the lengths in certain well-defined groups Or let us take a few rupee coins toss them and see how many times only one coin falls with head upwards how many times two fall in the same manner how many times three and so on In both these cases it will be found that the frequencies begin with small magnitude rise up to a certain maximum and begin to fall The former is a case of natural phenomena and the latter that of pure chance, and yet the rise and fall of frequencies would occur in identical manner in both of them

Let us take another example. Of the 111 students admitted to the B.A. class of a certain college in a particular year, 55 students were picked up at random their heights measured and tabulated as in table 47

Table 27. Heights of 55 boys arranged in ascending order.

Serial No.	Height in inches	Serial No.	Height in inches	Serial No.	Height in inches	Serial No.	Height in inches	Serial No.	Height in inches
1	59	12	63.5	23	66	34	67	45	68.5
2	60	13	64.5	24	66	35	67	46	68.5
3	61	14	64.5	25	66	36	67.5	47	69
4	61.5	15	64.5	26	66	37	67.5	48	69
5	62	16	65	27	66.5	38	67.5	49	69
6	62	17	65	28	66.5	39	67.5	50	69.5
7	62.5	18	65	29	66.5	40	68	51	69.5
8	62.5	19	65.5	30	66.5	41	68	52	70
9	63.5	20	65.5	31	67	42	68	53	70
10	63.5	21	65.5	32	67	43	68	54	71
11	63.5	22	65.5	33	67	44	68.5	55	72

From the table it will be found that (1) the height of 67 inches is repeated the largest number of times, (2) as we proceed on either side of 67 the number of times each height is repeated is not only less than the number of times 67 inches is repeated, but also the number of times each height occurs goes on falling, so that (3) a large majority of heights group round 67 inches. 67 inches is the modal height.

From these examples we can deduce the following law which would apply to most other cases: **Phenomena tend to fluctuate about a norm known as the mode, and a large majority of items cluster round it. As the distance from the mode widens, the items become fewer.**

We can arrive at this conclusion graphically as well We may plot the array of lengths of leaves or of the results of tossing coins We have plotted in Figure 22 the array of the height of 55 boys In this graph we see that near the extremes the heights change rapidly while the fluctuations are not so marked in the middle We further see that the mode is 67 inches since the largest number of lines stand for this number in the figure

Array of height of boys of 17 years chosen at random

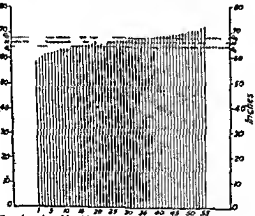


Fig 22

One might like to know whether the same result would follow if the heights of the remaining 56 students were also measured That is would the location of the mode be affected by the inclusion of more items? The answer is that if the 55 boys who were selected represented a fair sample of the whole class then the use of a larger number would give a greater regularity in the variation of the sizes and results not materially different from the one we have arrived at would follow

From Figure 22 it is very easy to locate the median and the two quartiles. The line drawn parallel to the base from the height of the 28th boy cuts the scale at 65.5 inches which is the required median. Similarly, quartiles can be located, as shown in the figure.

Frequency Graphs for Discrete Series.

The simplest mode of illustrating a discrete series is the **line** or **bar frequency diagram.** Size of item is taken

Line Frequency Diagram.

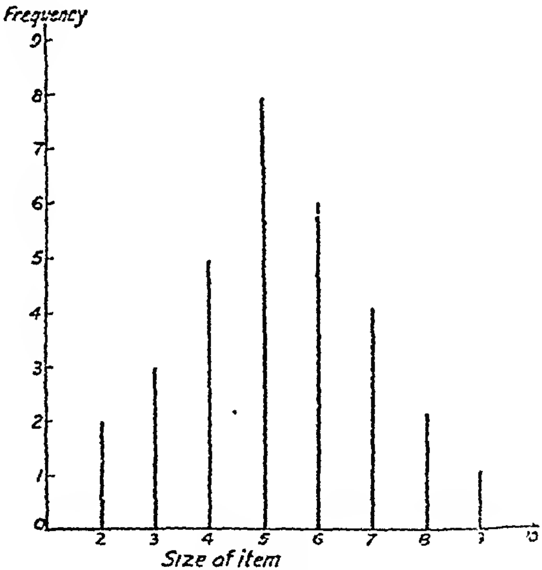


Fig. 23

on the horizontal line and the frequency on the vertical scale. Table 21, page 194, gives a discrete series which is graphically presented in Figure 23. Instead of drawing simple lines as done in the figure, bars of uniform thickness could also have been drawn to improve the appearance

of the figure The resulting figure would have been called *Bar Frequency Diagram* Sizes of items should be as they are in the figure separated on the horizontal scale by sufficient blank space so that the neighbouring lines may become absolutely distinct from each other

Frequency Graphs for Continuous Series Histogram

If in a series the range that is the difference between the largest and the smallest items is very large and instances occur at a great number of points between the two extremes the above method of the line or bar frequency diagram is not suitable to follow for it is impracticable to place a line at each measurement In such cases the data must be divided into classes and each class treated as a whole Table 48 gives such a classification of the data relating to the height of 55 boys given in table 47

Table 48 *Frequency distribution of the height of 55 boys*

Height in inches (size of item)	No of boys (frequency)
59—61	2
61—63	6
63—65	7
65—67	15
67—69	16
69—71	8
71—73	1

This frequency distribution can be illustrated by the **rectangular diagram or histogram** as shown in Figure 24 The histogram is composed of a set of rectangles one over each class interval on the horizontal scale The heights of the rectangles are in proportion to the frequencies in the class The area enclosed by the histogram is bounded by the lines of the

ordinates, the base line and the parallels to base line at the top of each class-interval.

Histogram representing the heights of 55 boys.

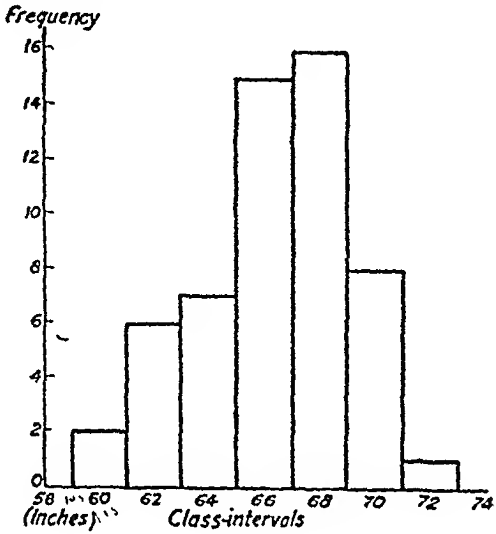


Fig. 24

The rectanguar histogram has a few characteristic features The series of rectangles in the figure illustrates fairly accurately the relative size of the various groups. The entire distribution of frequencies among the several classes becomes at once visible. The histogram is a better representative of the height of 55 boys actually measured than any smooth curve (Figure 25) would be, although the smooth curve would be a better representative of the heights of 111 students.

The total area of the rectangle erected on each class-interval is exactly equal to the number of frequencies in that class, the unit of area being measured by a rectangle one frequency unit in height and one class-interval in width. Thus the area of the figure equals the total number of

frequencies These two are, no doubt, significant facts of the rectangular histogram, but it is not without its defects

One defect of histogram is that different groupings would give different shapes If the class intervals in table 48 were made narrower, the steps in the histogram would decrease in size Secondly, it suggests, for instance, that there are 2 students each $\left(\frac{61+59}{2}\right)$, *i e*, 60 inches high, 6 students 62 inches high and so on As a matter of fact, each group consists of boys having different heights and therefore the rectangular presentation is misleading To do away with these defects a system of smoothing the histogram has to be devised so that a curve as typical of the entire data as possible may result For this purpose, Frequency Polygon or Frequency Curve may substitute the histogram

Frequency Polygon.

A simple method of smoothing is simply to connect the outer extremities of the base of the histogram with the mid-points of the tops of the rectangles, as is shown by the dotted lines in Figure 25 In the figure, the lines connecting the mid points of the tops of the rectangles have been extended to the base at points 58 and 74 inches, the mid points of the two rectangles outside the histogram at which the observed frequencies are zero This procedure gives an area representation of the frequency distribution which is exactly equal to the area of the histogram The triangular strips of area which are excluded from the histogram are equal to those formerly outside the histogram but now included in the polygon. (Compare a and a', b and b', c and c', etc) Thus the area of the polygon is equal to the area of the histogram But the area of each rectangle of the histogram is not

equal to the area of the corresponding section of the polygon. For, the area cut off from the class-interval 61-63, say, has been added to the preceding class, 59-61. To this extent, the polygon may be said to have re-distributed the frequency distribution.

The Frequency Polygon and the Frequency Curve representing the heights of 55 boys.

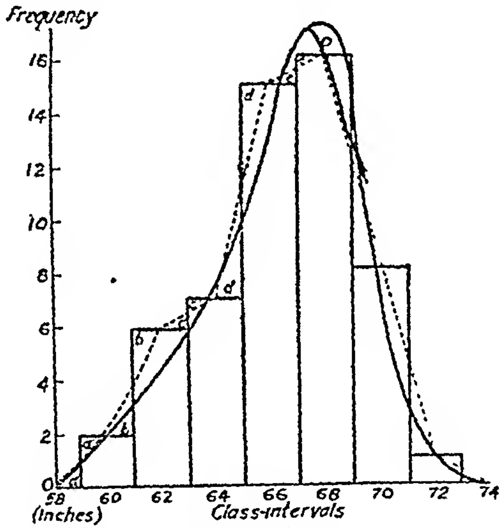


Fig. 25

The main purpose of the polygon is to find the mode of the given series. Mode can be ascertained fairly accurately by the apex of the polygon. Apex would in all probability occur in the class-interval containing the mode, and will not be shifted greatly even if some more items are added to the series so that the original frequency distribution is modified.

The frequency polygon, however, has a defect. It shows sudden changes in the direction of the curve, particularly at the apex, P. It, therefore falls to show regular and uniform variation in magnitude, which purpose is well served by a continuous smooth curve, called Frequency Curve.

Frequency Curve

Frequency polygon gives the first approximation to making a continuous smooth curve In Figure 25 frequency curve also called **smoothed histogram** has been drawn free-hand and is shown by the continuous line Smoothing free-hand requires great care which if not taken would lead to fallacious presentation of facts When smoothing a frequency polygon the fact that it is really derived from the histogram should always be kept in view This would imply that the top of the curve would overtop the highest point of the polygon particularly when the magnitude of class intervals is large Again the curve should look as regular as possible all sudden turns should be avoided The extent of smoothing permissible would however depend on the particular data under study If the data consist of records of natural phenomenon like the measurement of leaves or of chance phenomenon as the tossing of rupee coins smoothing may freely resorted to since such phenomenon under study is social or economic skewness sometimes considerably large may be expected in the normal curve In smoothing such a polygon only minor irregularities may be eliminated The smoothed histogram should begin and end on the base line since a continuous series which it represents begins with very few instances which go on rising but decrease again slowly to zero As a general rule it may be extended to the mid point of the class intervals just outside the histogram Another fact that must be kept in view while smoothing a frequency polygon is that the area under the curve should represent the total number of frequencies in the entire distribution In the matter of smoothing experience is the best teacher

Frequency curve has certain characteristics. In most cases particularly in natural and chance phenomena it is

bell-shaped. Bell-shaped curve is also called the **Normal Frequency Curve** or the **Normal Curve of Error.** Normal curve indicates what is expected of the phenomenon to which the curve relates under normal conditions. This curve eliminates accidental variations and establishes normal tendencies. If such a curve has been once obtained with adequately representative sample, it can be utilized to speak for the whole universe. For instance, it may be said that if more measurements are taken, not only will they fall within the curve, but most of them would be found to cluster round the mode. Or, if the groups are re-arranged, those groups which are nearer the modal group will contain, as a rule, more cases than the groups more distant.

To draw the frequency curve it is necessary first to draw the polygon. The polygon is later smoothed out. Frequency polygon may be drawn, even without first drawing the histogram, by plotting the frequencies at the mid-points of the class-intervals and joining them by straight lines. This is no doubt easy but presents difficulties in smoothing the polygon properly. Therefore, it is always safe to proceed in a sequence —first draw the histogram, then the polygon and lastly smooth it keeping in view the fact that the area of the curve should equal that of the histogram.

Ogive Curve.

Of the three methods of presenting frequency distribution —the histogram, the frequency polygon, and the frequency curve—the last is the best for many purposes. But these methods are based on the frequencies of the class-intervals and not on the cumulative frequencies. Ogive curve is based on cumulative frequencies and is, therefore, also designated as **cumulative frequency curve.**

Table 49 gives the frequency distribution of 204 weekly market quotations of the shares of the Imperial Bank of India

taken week by week from 1st January 1933 to 22nd December 1936

Table 49 *Cumulative Frequency Table showing market values of the shares of the Imperial Bank of India (Paid up Value Rs 500)*

Market value of share	No of quotations (frequency)	Cumulative frequency
1150 — 1200	11	11
1200 — 1250	44	55
1250 — 1300	9	64
1300 — 1350	10	74
1350 — 1400	7	81
1400 — 1450	6	87
1450 — 1500	12	99
1500 — 1550	30	129
1550 — 1600	51	180
1600 — 1650	20	200
1650 — 1700	4	204

To construct cumulative frequency curve from the table the horizontal and the vertical scales are chosen in the same manner as in the case of histogram polygon or curve but the essential difference between the plotting of frequency polygon and of ogive curve is that in the polygon the frequency must be plotted at the mid point of the class interval while in the ogive it must always be plotted at the upper limit of the class interval Thus in Figure 26 we mark 11 against Rs. 1,200 55 against Rs 1,250 and so on The successive points are later connected by straight lines with a ruler The resulting curve is an ogive curve. This curve can also be smoothed much like the smoothing of frequency polygon but it has not been done in Figure 26.

Ogive curves or simply ogives may be used for the purpose of comparing groups of statistics in which time is

not a factor. Ogives, in general, are not easy for the ordinary person to interpret. Histograms are readily understood by him. Ogives are primarily drawn for determining medians, quartiles, percentiles etc. To determine the median of the data given in table 49, a line from the mark $\left(\frac{204}{2}\right)$ or 102 on

Ogive Curve of The Frequency Distribution of the values of 204 shares of the Imperial Bank of India

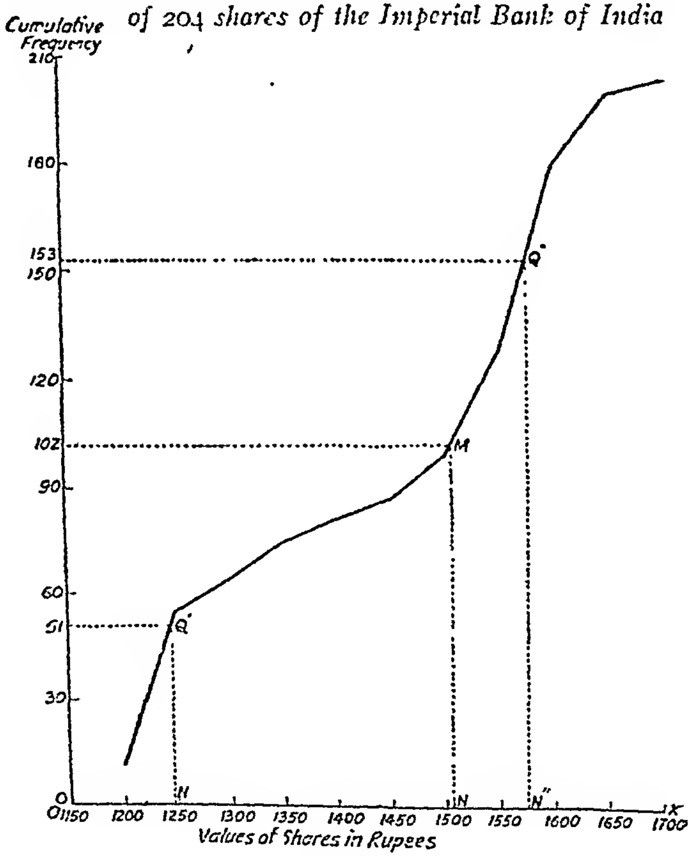


Fig. 26

the vertical scale is drawn parallel to the horizontal axis to intersect the ogive at M (see Figure 26), and then a perpendicular

is drawn from M on OX cutting it at N. ON, read from the figure gives the median which is Rs 1,305. Similarly ON gives the value of the upper quartile as Rs 1,376 and ON the value of the lower-quartile as Rs 1,245. Deciles and percentiles can also be likewise determined. This method of locating the median etc. is far more easy and simple than the methods discussed in Chapter X and is particularly so when the data given are imperfect.

The ogive is useful for yielding other results as well. Suppose an ogive represents the cumulative frequencies of income-tax payers in a certain country. We can from it, easily find out the total number of tax payers paying not less than a certain sum. Again if data relate to wages of employees in a factory, the number of workers getting not less than a certain wage can be ascertained. Similarly if the government of a country wishes to formulate a scheme of graded retrenchment this method of determining the number of employees getting not less than a certain salary would be found very useful. For simply an ordinate need be drawn from the amount of money under consideration in the three cases to intersect the ogive and the value of this ordinate be read on the vertical scale to know the number of tax payers, wage-earners or government employees as the case may be. Further the mode can also be located on the ogive as the frequencies are most numerous where the curve has the greatest tendency to run parallel to the vertical scale.

Galton's Method of Locating the Median[1]

Galton has given a graphic method by which median can be located. The horizontal line is divided into equal parts corresponding to the unit of measurement and vertical line is similarly divided to show the frequency. The essential point

[1] It should not be confused with Galton Graph.

to be kept in view while plotting the curve is that every preceding measurement is made the base for the next measurement.

Table 50 reproduces the data given in table 47, page 359, in the form of frequency distribution.

Table 50. *Frequency Distribution of heights of 55 boys.*

Height	Frequency	Height	Frequency	Height	Frequency
59	1	64.5	3	68	4
60	1	65	3	68.5	3
61	1	65.5	4	69	3
61.5	1	66	4	69.5	2
62	2	66.5	4	70	2
62.5	2	67	5	71	1
63.5	4	67.5	4	72	1

Locating the Median by Galton's method in a series of heights of 55 boys.

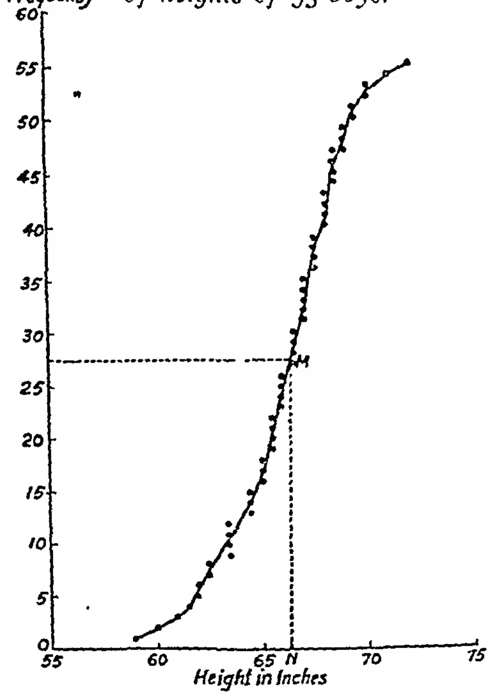


Fig. 27

The frequency distribution given in table 50 is graphically represented by Galton's method in Figure 27 Starting with 59 one dot is put down on the ordinate through the 59th mark standing for one student From his point a horizontal line is drawn upto 60 the next height on the ordinate through the point 60 Starting from the new base only one dot is again marked From this dot again a horizontal line is drawn and proceeding in the same manner as many dots are marked at each successive height as there are students of that height Thus 55 dots are put down When this has been done lines are drawn to connect every two successive dots which are horizontally apart Where the dots are odd in number, the line passes through the middle dot while the line passes through the point midway between the two middle dots if the number of dots is even. Thus a continuous curve is obtained To locate the median the position of 27½th point is marked on the vertical scale since the height of the 27th boy is the median From this point a horizontal line is drawn to intersect the curve From the point of intersection a perpendicular is drawn on the horizontal base to intersect it at N ON gives the value we desire Thus 66 5 inches is the median of the heights of 55 boys

EXERCISES

(1) The wandering of a line is more powerful in its effect on the mind than a tabulated statement Elucidate

(2) What points must be borne in mind in drawing a statistical graph?

(3) Explain with the help of diagrams the difference between a frequency polygon histogram frequency curve and ogive

(M.A Patna 1942)

(4) How will you compare the proportional changes in two or more variables? Give the details of the procedure you would adopt

(5) How does the Natural Scale differ from the Ratio Scale? In which cases should the latter scale be used?

(6) Describe the Galton's method of locating the median.

(7) The following table gives the value of Imports and of Exports of India for the years 1920—21 and 1921—22 in crores of Rupees:

Months	1920-21		1921-22	
	Imports	Exports	Imports	Exports
April ..	22	28	26	18
May ..	24	28	21	20
June ..	26	23	19	17
July ..	28	21	18	17
August ..	31	20	21	20
September ..	29	22	20	20
October ..	32	21	23	18
November ..	32	19	26	20
December ..	32	20	23	22
January ..	31	19	28	23
February ..	24	18	20	22
March ..	25	19	21	28

Plot the above figures on a graph paper, and show also the balance of trade.

(B. Com., Alld., 1938).

(8) Following are the monthly cheque clearances in India during 1942-43. Present them graphically and give necessary comments.

Months			Cheque Clearances Crores of Rs.
April	..	..	181.5
May	..	..	219.6
June	..	..	188.2
July	..	..	193.0
August	..	..	244.6

Months	Cheque Clearances Crores of Rs
September	2130
October	2366
November	2330
December	2386
January	2627
February	2212
March	2138

(9) What are the advantages of the Ratio Scale over the Natural Scale? Plot the following data graphically on the logarithmic scale —

Year	Total Notes issued in crores of Rs	Notes in circulation in crores of Rs
1933 34	177	167
1934 35	186	172
1935 36	196	167
1936-37	208	192
1937 38	214	185
1938 39	207	187
1939-40	252	237
1940-41	269	258
1941 42	421	410
1942 43	650	625

(B Com. Nagpur 1943)

(10) The following table shows the total sales of gold by the Bank of England on foreign account Represent the data graphically on the logarithmic scale —

Year	£000
1910	14,498
1911	8,228
1912	9 670
1913	7 943
1914	8 077
1915	43 076
1916	[illegible] 360

(B Com Alld. 1932)

(11) Represent graphically the data given below on a single sheet of graph paper to bring out clearly the relative fluctuations in the prices of various articles Draw such conclusions as you can from the graphs

Wholesale prices in Kanpur.
(in rupees per maund)

Year		Rice	Wheat	Linseed	Gur	Cotton	Tobacco
1928	..	7.3	7.7	7.0	6.5	34.1	17.3
1929	..	7.7	5.5	8.0	7.3	29.8	17.1
1930	..	5.8	3.6	6.5	6.2	17.3	14.5
1931	..	4.1	2.7	4.2	4.2	13.3	11.6
1932	..	4.3	3.4	3.5	3.5	14.8	4.9
1933	..	4.1	3.2	3.4	3.1	12.9	4.9
1934	..	3.7	2.8	3.6	4.1	13.2	5.7

(M. Com., Alld., 1943).

(12) Show the results of working of Class I railways in India graphically and comment thereon.

(In millions of £)

			Capital outlay	Gross Earnings
1923-24	..	..	464	70
1924-25	..	..	473	74
1925-26	..	..	487	73
1926-27	..	..	505	72
1927-28	..	..	594	86
1928-29	..	..	599	86
1929-30	..	..	617	84
1930-31	..	..	627	77
1931-32	..	..	631	71
1932-33	..	..	638	70
1933-34	..	..	635	72

(B. Com., Agra, 1940).

(13) Distribution of firms in Woollen and Worsted Industries in Yorkshire, according to number of operative:

Operatives	No. of firms	Operatives	No. of firms
1— 20	380	301—340	24
21— 60	320	341—380	18
61—100	182	381—420	16
101—140	147	421—460	11
141—180	92	461—500	9
181—220	66	501—700	19
221—260	39	701—900	15
261—300	30	901—	16
Total Number of firms			1384

Represent this distribution graphically (by means of a cumulative diagram) and from this graph estimate the median and quartiles of the group.

(B Com Luck 1930)

(14) Present the following figures relating to monthly imports (volume and value) of liquor into India graphically so as to show their fluctuations and give necessary comments.

1941-42 Months	Volume (000) Gals	Value Rs (000)
April	163	2649
May	395	193
June	358	2113
July	415	2655
August	330	2104
September	363	2339
October	456	3105
November	349	2310
December	209	1526
January	230	2107
February	159	1667
March	35[illegible]	1993

(15) Plot the following figures relating to population of India so as to show the proportionate increase in population from one period to another.

Year	Population (000,000's omitted)
1872	10
1881	50
1891	290
1901	295
1911	315
1921	3[illegible]
1931	350
1941	390

(B Com Nag [illegible])

(16) Graphically present the figures given in exercise 24, Chapter X, page 175, and state whether the curve is skew. If yes, what is the nature of skewness—positive or negative?

(17) Draw a line frequency diagram from the figures given in exercise 31, page 222, Chapter XI.

(18) Study the movement of the exports of pig iron and of cotton goods from the figures given in exercise 26, Chapter XI.

(19) Draw a bar frequency diagram of the figures given in exercise 28, Chapter XI, page 221.

(20) Draw a frequency curve from the data given in exercise 21, Chapter XI, page 218.

(21) Plot the figures regarding area under cultivation, production and price of wheat as given in exercise 33, Chapter XI, page 222, on a graph paper, and comment upon their relationship. (M.A., Patna, 1941).

(22) The following table gives the age distribution of widows in India (Census Report 1931). Draw a graph showing the number of widows younger than any given age, and from the graph read off the median age of the widows and also the upper and the lower quartiles.

Years			No. of widows
0—10	..	..	135,862
10—20	..	..	718,101
20—30	..	..	2,456,835
30—40	..	..	4,847,631
40—50	..	..	6,480,259
50—60	..	..	5,908,159
60—70	..	..	3,743,615
70 and over	..	..	1,957,506
	Total	..	26,247,968

(M.A Alld, 1942).

(23) Locate the median of the following figures by Calton's Method —

Length of Nim leaves in inches

1.35, 1 35, 1.6, 1.6, 1.7, 1.7, 1.9, 1 6, 1.5, 1 9, 2.0, 2.3, 2.6, 2.8, 2.5, 2.3, 2.9, 3.4, 3.7, 2.9, 3.2, 3 4, 2 5, 2 8, 2 8, 2.6, 2.5, 2.3, 2 4, 2 7, 2.7, 2.7, 2.9, 3 3, 1.8, 1.6, 1.5, 1 9, 1.5, 1.6, 3.4, 2.7, 3.9, 3.5, 2.9, 2 1, 2.2, 2.3.

(24) In the following table are given the quantity of white (bleached) cotton cloth imported into India and the price per yard Bring out, graphically, the relation between price and quantity imported year by year, and comment on the relation indicated.

Years	Cotton cloth imported million yds	Average price per yard Rs a p
1924—25	549	0 6 0
26	465	0 5 6
—27	571	0 5 0
—28	557	0 5 0
—29	554	0 4 6
—30	474	0 4 6
—31	27[illegible]	0 3 9
—32	[illegible]80	0 3 0
—33	412	0 3 0

(25) Following table gives the production of sugar in Cuba Java and India during 1930-1939 in millions of quintals Represent the figures by a suitable diagram and comment on their relationship

Year	Cuba	Java	India
1929—30	44	29	17
1930—31	30	28	20
1931—32	25	26	24
1932—33	19	14	28
1933—34	2[illegible]	6	30
1934—35	25	5	31
1935—36	25	6	36
1936—37	29	14	40
1937—38	29	14	32
1938—39	26	15	27

(M A Patna 1943)

(26) The following table gives the prices of gold and wheat and net export of gold during the years 1931 32 to 1938 39 —

Years	Average price of gold (per tola) Rs a	Average price of wheat (per maund) Rs	Net export of gold Crores of Rs
1931 32	25 4	3 3	58
1932 33	30 12	3 3	65
1933 34	33 6	2.8	57
1934 35	35 8	3 1	52
1935 36	35 4	3 2	37
1936-37	36 0	3 9	28
1937-38	36 6	3.0	16
1938-39	37 12	3.4	13

Plot the above figures on a graph paper, and comment upon the relationship

(M A Agra 1943)

CHAPTER XVI

ANALYSIS OF TIME SERIES

In examining the changes with time of a certain quantity we are concerned with the interpretation of these changes and with finding how they are related to similar changes which are observed in other time series. For instance, when we examine the series of index numbers giving the relative fluctuations of retail price of wheat in India (1873=100). given in column 2 of table 51 on the next page, we naturally ask ourselves to what the changes taking place are due and how they are related to changes in other series.

Trend, Seasonal and Cyclical Fluctuations.

Upon perusal of the annual average indices we find that there is, on the whole, a gradual increase in the retail price of wheat and that there are sudden breaks of large and small number of points in this gradual increase. We know that the series of retail price is a resultant of a large number of causes of different kinds, *e.g.*, weather conditions, transport facilities, consumers' demand for wheat, demand and prices of substitutes of wheat and so on. We must consider the nature of these causes with a view to determine their effects.

The first cause, and a fundamental one, is that with increase in population of India the demand for wheat, as for so many other commodities, has been growing. Therefore, there is a certain *growth factor* which is effecting a general and gradual rise in the series. The resulting gradually changing nature of the retail price of wheat is referred to as the **secular trend** of the series and this trend should be supposed to be linked up with the growth factor.

Table 51 *Index Numbers of Retail Price of Wheat (1873—100) and Trend by the Method of Moving Averages*

1 Year	2 Annual average	3 5 Yearly moving average	4 10 Yearly moving average	5 10 Yearly moving average (centred)	6 Deviation from moving average	1 Year	2 Annual average	3 5 Yearly moving average	4 10 Yearly moving average	5 10 Yearly moving average (centred)	6 Deviation from moving average
1873	100					2	151	1[illegible]7	135	131	+20
4	94					3	127	127	137	136	— 9
5	81	91				4	103	130	140	139	—31
6	[illegible]8	100				5	116	141	145	143	—27
7	102	113				6	153	145	148	147	+ 6
8	14[illegible]	121	103	108	+39	7	[illegible]06	152	147	148	+58
9	158	124	105	1[illegible]8	+50	8	143	164	147	147	— 4
1880	118	124	108	108	+10	9	140	163	149	148	— 8
1	96	113	111	110	—14	1900	170	152	151	150	+20
2	101	102	113	112	—11	1	160	149	152	152	+ 8
3	103	96	110	112	— 9	2	142	146	148	150	— 9
4	91	97	10[illegible]	108	—17	3	1[illegible]9	139	156	153	—23
5	89	101	106	106	17	4	123	138	162	159	—31
6	103	10[illegible]	110	108	— 5	5	142	143	162	162	—20
7	1[illegible]1	111	11[illegible]	113	+ 8	6	15[illegible]	163			
8	1[illegible]9	110	113	11[illegible]	+ 6	7	168	1[illegible]9			
9	118	1[illegible]8	119	119	— 1	8	[illegible]96	184			
1890	11[illegible]	1[illegible]1	1[illegible]	121	— 4	9	203				
1	[illegible]3	1[illegible]0	12[illegible]	1[illegible]	+12	1910	1[illegible]0				

Just as the growth factor causes a rising tendency in the indices of retail price of wheat in India, the *decline factor* in some other series, say, production of indigo in India over a series of years would effect a falling tendency. The downward trend in such a series would then be said to be linked up with the decline factor.

Secondly, operating along with the growth or decline factor is a group of causes which do not operate continuously, but in a regular spasmodic manner. One among these causes is the *seasonal factor*. Seasons occur in the same way every year: Winter is followed by summer and summer by rains in India; crops have their sowing and harvesting seasons; May and June constitute marriage season in India. The effect of this seasonal factor is a regular up and down movement in the series of figures relating to the phenomenon affected by the factor. This movement is referred to as the **seasonal movement.** If retail prices or retail price indices week by week or month by month are considered, an up and down movement of this kind would be noted due to the harvesting season in March-April, and this movement would be super-imposed on the secular trend.

Another cause in this group, operating in a regular spasmodic manner, is the *cyclical factor*. During the 19th century a fairly regular up and down movement has been noticed in a good number of time series of economic data These movements have been repeated at intervals of years ranging from 7 to 11, that is, these movements have occurred in a cycle. There are "boom" years in which the observed phenomenon shows upward movement, and there are "depression" or "crisis" years when it shows downward movement. Retail price of wheat is also affected by this "trade cycle" or **cyclical movement.**

Lastly, in addition to the group of causes producing regular up and down movement in our series, there is an-

other group which operates in an *irregular* manner It includes such events as floods raids of locusts fires earthquakes wars revolutions and so on which ruin the crops of the areas affected by them It also includes such chance combinations of wind sunshine and rain in a certain season as may result in a bumper crop While these causes do operate from time to time there is no regularity in their operation Retail price of wheat is also affected by such **Irregular fluctuations**

We may conclude then that in analysing any given time series we look for three kinds of movement viz

1 General Trend or Secular trend
2 Regular Fluctuation
   (i) Seasonal
   (ii) Cyclical
3 Irregular Fluctuations

General trend refers to secular or **long period** changes Some influences operating steadily and persistently from year to year may be causing a general tendency for figures relating to a certain phenomenon to increase to decrease or to assume both directions Regular and irregular fluctuations refer to periodic changes lasting a **short period** of time Therefore changes in time series may be spoken of as (1) long time and (2) short time Long time fluctuations include secular changes while short time fluctuations may be classified into (i) seasonal (ii) cyclical and (iii) irregular fluctuations The primary task in analysing a time series therefore is to measure and isolate long time and short time changes

Measuring and Isolating Time Changes.

In order to study any one of the changes by itself it seems necessary to follow the method of the physicist who allows only one factor to vary at a time and eliminates all other factors But the statistician can rarely control the condition

of his experiment and has, therefore, to be satisfied with ridding, so far as possible, the recorded data of the apparent effects of the extraneous causes. If we desire to study the long-time changes in the retail price of wheat, we shall do well to remove the short-time fluctuations from the field. But, if we are interested in short-time oscillations only, we should eliminate all long-time changes from our series.

Let us first be concerned with the study of long-time variations, or which comes to the same thing, trend. Since *the value of an item on a particular date in a time series consists of the long-time plus the short-time changes,* we can get a measure of the long-time change if we eliminate short-time change from the series.

Measurement of Secular Trend.

If we plot our series on a graph paper we shall observe an up and down movement in the curve. The index histori-gram of the retail prices of wheat drawn in Figure 28, page 386, is not a smooth curve. It is irregular. If we can smooth out these irregularities, the short-time oscillations shall be removed, and we shall obtain a measure of the trend values. Use may be made of any one of the following methods for measuring long-time or secular trend, and, in the process, eliminating short-time oscillations from the data:

(1) Freehand Curve Method.
(2) Method of Moving Averages.
(3) Method of Least-Squares.

Freehand Curve Method.—We may observe the up and down movement of the curve and smooth out the irregularities by drawing a freehand curve or line through the index histori-gram such that the curve so drawn would give a general notion of the direction of the change. This freehand curve eliminates the short-time oscillations and shows the long

period general tendency of the retail price of wheat This is exactly what is meant by trend

But this method has a serious disadvantage that different people would draw the freehand line at different positions with different slopes Naturally there will be different conclusions Therefore in place of this method the method of moving averages may be used

Method of Moving Averages.—It is an alternative method of ridding the historigram of its fluctuations It involves the taking for each year of the series not the value relating to that year but the average of the values of one two, three or more years preceding and succeeding the year in question If for instance three-yearly moving average is to be computed, the values of the 1st 2nd and 3rd years are added up the sum is divided by 3 and the quotient is placed against the 2nd year, then values of the 2nd 3rd and 4th years are added up averaged and the average is placed against the 3rd year, and so on These averages when plotted on the same graph on which the historigram has been drawn would smooth out its irregularity, that is eliminate short time changes and show the long period tendency

What period of time should be used in calculating the moving average? This question is of paramount importance since, no one period of time used in calculating the moving averages would eliminate short time changes of all time series The moving averages calculated by using that time-interval which approximately coincides with the periodicity of a historigram would eliminate, *nearly* completely, all regular fluctuations of the series and show the trend. This is the chief characteristic, as also the merit of the Method of Moving Averages In using this method therefore, we must search for 'periodicity' in the given series to

determine the period of time that should be utilised in computing the moving averages.

What is 'periodicity' and how is it measured? Periodicity is but another name for the average duration for which cycles in a series last. A cycle is said to be completed when beginning with a peak the falling curve reaches a minimum point and then rising again reaches the next peak, or, when beginning with a trough the rising curve reaches a maximum point and then falling again reaches the next trough. There may be several peaks and troughs in a given series. The simple average of the time-distances between the consecutive peaks and between the successive troughs of the waves of the historigram would give the approximate wave-length, the desired periodicity.

In Figure 28 the index-historigram is undoubtedly irregular in its shape, but the ups and downs show remarkable regularity in their occurrence. It will be observed that it shows recurring years of high prices at intervals of about 10 years, prominent crests falling in the years 1879, 1888, 1897 and 1908, and similarly recurring years of low prices again at intervals of about 10 years, prominent troughs falling in the years 1876, 1885, 1894 and 1904. The average of these time-distances is 10 years, and it may now be said that the 'periodicity' of the given series is ten years, that is, the ups and downs repeat themselves in a cycle of nearly 10 years with marked regularity. These regular fluctuations of the index-historigram are the *Cyclical fluctuations*, or, call them *periodic variations*. If we eliminate them, we shall be left with the long-period trend. To do so, we should compute ten-yearly moving averages of the retail price indices, and plot them over the index-historigram. The ten-yearly moving average curve would show the secular trend.

To be more certain whether the regular fluctuations repeat themselves every tenth year, we operate on the series with moving

averages of a few different periods and observe if any other moving average smooths out the irregularities. We begin with

Retail price of wheat in India showing annual fluctuations, five yearly moving average and the trend

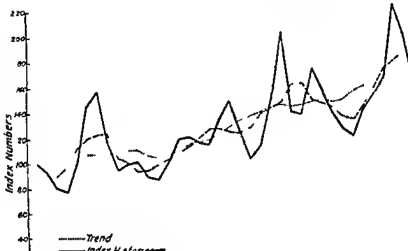


Fig. 98

five-yearly moving average which is given in column 3, table 51. This series of moving averages is plotted over the histogram in Figure 98. The five yearly moving average curve shows considerable fluctuations, though they are less than those in the histogram of annual indices. Since it does not smooth out the irregularities, it cannot be regarded as showing the secular trend of price indices. But a careful glance at the successive peaks and troughs of this curve would reveal a periodicity of 10 years for the given series. If we similarly plot seven-or

nine-yearly figures, we would even then find some fluctuations in the moving average curves, until we come to the ten-yearly moving average.

Ten is an even number and, therefore, the ten-yearly moving average can be placed only against the middle of the fifth and the sixth years of each group, as in column 4 of table 51. We 'centre' these ten-yearly moving averages by taking two-yearly moving average of the figures given in column 4. For example, the ten-yearly moving average (centred), shown in column 5 of the table, for 1881 is

$\frac{108+111}{2}=110$, and for 1887 is $\frac{110+115}{2}=113$.

If, however, the periodicity were an odd number of years, there would have been no necessity to 'centre' the moving averages. The curve of the ten-yearly moving average (centred), drawn through the historigram in Figure 28, is far more smooth than the annual index-historigram or the five-yearly moving average curve. This curve shows the general rising tendency of wheat prices from 1873 to 1910. Consequently, the ten-yearly moving average (centred) column in table 51 gives us the trend values. *We have thus measured the trend by eliminating short-time fluctuations through the process of smoothing.*

The smoothed curve is of no use for studying short-time changes. It shows the trend, the course that would be taken by a curve in the absence of disturbing factors, such as cyclical fluctuations. From the smoothed curve in Figure 28 we cannot study when prices began to rise or to fall. We cannot say that the retail price of wheat began to rise in 1887 because the smoothed curve begins to rise in that year The average for 1887 is based on the prices of ten years of which the year in question is only one. *To study short-time (annual) changes, that is, to study when prices began to rise and to fall, we*

must consult the original historigram and not the trend If we study them from the trend misleading conclusions might be drawn. For example the shape of the smoothed curve for the period 1878 to 1880 (Figure 28) might lead one to think that the price of wheat was fairly steady during that period whereas according to the original data the indices 147 158 and 118 for the years 1878 1879 and 1880 respectively show violent fluctuations The smoothed curve therefore is good only for a study of long period general tendency of the phenomenon under investigation

The Method of Moving Averages has two limitations in spite of its being easy of application and enjoying an advantage over the freehand curve method that different people will not obtain different results by using it provided of course, the given figures show periodic variations

The first limitation is that it does not enable the carrying out of the accurate trend to the extremes of the data Trend values relating to the years from 1873 to 1877 and to those from 1906 to 1910 could not be ascertained in our example This deficiency may be made good in either of the two following ways

(a) The moving average curve may be carried out freehand to each extreme

(b) Artificial final groups may be formed by duplicating the numbers at the extremes In table 51 for instance 170 might be added at the close five successive times forming the required new groups for computing 10-yearly moving average upto 1910

Both these methods are however approximations

The second limitation of the method of moving averages is that it cannot be applied with equal success to any and every historigram It is useful only in those historigrams which manifest more or less periodicity for the object of using

this method is to eliminate periodicity. If a historigram does not show regular periodicity, the period of the moving average to smooth out its irregularities would obviously be very long, and the moving average would show the general trend for the whole period without allowing for any of the large variations which it might be proper to retain Hence, the method of least-squares.

Method of Least-Squares.—Although somewhat involved for practical use this is the most representative method. By applying it, a straight line trend can be fitted to the gven data The trend line, technically called the line of "best fit," drawn under the least square assumption is the line from which the sum of the squares of the items, measured parallel to the y-axis, is the least. The following table illustrates the procedure to be adopted:

Table 51A. *Computation of Trend of the Sterling Assets of the Reserve Bank of India by the Method of Least-Squares.*

Year (1)	Sterling assets (crores of Rs) (2)	Time deviation from middle year (3)	Deviation of col. 3 squared (4)	Product of col 2 and col. 3 (5)	Trend ordinates (crores of Rs) (6)
1936-37	83	−2	4	−166	67
1937-38	92	−1	1	− 92	84
1938-39	71	0	0	0	101
1939-40	90	+1	1	+ 90	118
1940-41	169	+2	4	+338	135
	505		10	+170	

The procedure may be explained as follows:

(1) Compute the simple arithmetic average of the items. It is 505÷5=101 crores of Rupees. This average is the mid-point of the line of "best fit."

(2) From the mid-year of the series, 1938-59, and the time-deviations of the other years, Col. (3).

(3) Square the deviations and add the resulting figures Col (4)

(4) Multiply the value of the items by the time-deviations and take the algebraic sum of the products Col (5)

(5) Divide the sum of Col (5) by that of Col (4) The quotient Rs 17 crores is the average increment in the trend year by year It gives the slope of the line of best fit

(6) To obtain the trend ordinates shown in Col (6) put down the arithmetic average of the items Rs 101 crores against the mid year 1938 39 and for the years after the mid year add as many increments to the trend ordinate of the mid year as the given year is away from the mid year while for the years before the mid year subtract the increments in a similar manner Thus the trend ordinates for 1939 40 and 1940-41 are respectively (1×Rs. 17 crores) = Rs 17 crores and (2×Rs 17 crores) Rs 34 crores more than the average while the trend ordinates for 1937 38 and 1936-37 are respectively (1×Rs 17 crores) −Rs 17 crores and (2×Rs 17 crores) −Rs 34 crores below the average The algebraic sign of the increment must be kept in view while computing trend values

The values of Col (6) when plotted would show the straight line trend of the sterling assets of the Reserve Bank of India It should be clearly noted that this trend line the line of best fit passes through the arithmetic means of the x and the series x series representing time and y the sterling assets This line it should be further noted is the best fit only for the period to which it has reference the addition of sterling assets for a few more years or the elimination of some of those already included may change the position materially

Elimination of Trend.

We are very often interested in the study of short time oscillations of a time series For this study we should get our data rid of long time variations To do so is quite easy

when once the trend of the series has been known. for, the difference between the value of an item on a particular date and its corresponding trend value is the short-time oscillation. These differences, called deviations, are taken wth their proper algebraic signs, and, when plotted on a horizontal base line show the short-time oscillations.

We have discovered the trend of our series relating to retail price of wheat in India. The values of the ten-yearly

Retail price of wheat in India showing short-time oscillation (with the trend removed).

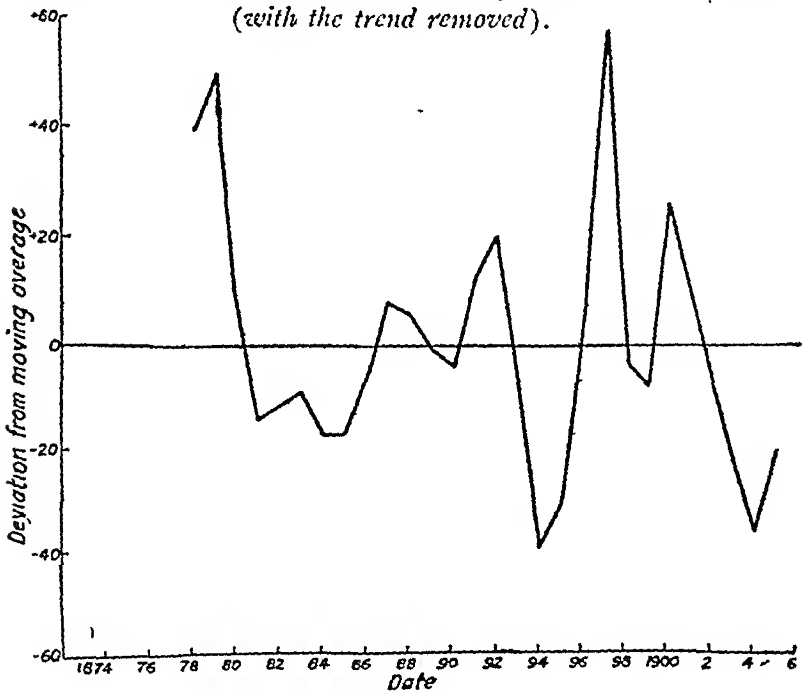


Fig. 29

moving averages in column 5, table 51, are the trend values. To eliminate the trend we compute deviations of the prices from the trend values. These deviations are placed in column 6 of the sáme table. We plot them on a graph in Figure 29.

The resulting curve gives only the short time oscillations of retail price of wheat unobscured by the long time variations Fluctuations in price without the trend can now be studied from the diagram very easily

Thus the primary task of studying and eliminating long time and short time changes in a time series is over The short time oscillations would it may be observed consist of seasonal cyclical and irregular fluctuations which may also be separately measured

Measurement of Normal Seasonal Change

If monthly records of some data are available over a series of years either of the following two methods may be used to arrive at a measure of their normal seasonal variation

(1) Computation of Seasonal Average
(2) Computation of Seasonal Index by using monthly averages

Seasonal Average—Seasonal average is obtained by adding up the amounts separately for each month of the year, that is totalling the amounts for the Januaries the Februaries etc. separately and dividing each such summation by the total number of years for which the data are given In table 52 column 8 gives the seasonal average for the exports of raw jute from India between 1939 and 1943

Seasonal Index.—To obtain the seasonal variation index find out the average of monthly averages and then find the per centage that each monthly average bears to the average of monthly averages In table 52 the average of monthly averages given at the bottom of column 8 is 28000 tons Therefore the seasonal index for January is

$$\frac{42000\times100}{28000}=150$$

Seasonal indices in column (9) of the table have been similarly calculated

Table 52. *Calculation of Seasonal Average and Seasonal Variation Index of the Exports of Raw Jute from India.*

Months	Raw Jute					Total of five years	Five years' average	Percentage of Monthly total to average Monthly total
	1939	1940	1941	1942	1943			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Tons (000)	Tons (000)	Tons (000)	Tons (000)	Tons 000)	Tons 000)	Tons (000.	
Jan.	62	59	45	33	22	211	42	150.0
Feb.	69	68	17	15	14	183	37	132.2
March	60	51	10	24	13	158	32	114.3
April	53	38	20	26	17	154	31	110.7
May	44	36	31	6	12	129	26	92.9
June	34	16	36	16	12	114	23	82.0
July	21	8	26	28	13	96	19	67.9
Aug.	37	14	30	14	8	100	20	71.4
Sept.	45	11	23	22	11	112	22	78.6
Oct.	31	17	29	21	15	113	22	78.6
Nov.	65	12	30	38	22	167	33	117.8
Dec.	60	35	16	22	13	143	29	103.6
Total						1680	336	1200.0
Average						140	28	100

The seasonal average in column (8) and the seasonal index in column (9) show the typical movement of exports of raw jute. If both of them are plotted on graphs, the part of the year in which the exports are the greatest and the part in which they are the least would become easily visible. The influence of the seasonal factor on the exports would thus be rendered clear. The normal seasonal variation, thus computed, would recur month by month in the same manner year after year.

In table 52 we have pooled together the experience of only five years. In practice a longer period should be used.

Comparison of Time Changes in two Historigrams.

When comparison of time changes in *two* historical variables is desired, they should first be reduced to index

numbers so that their relative fluctuations may be easily compared. The trends of the two index series may then be discovered their short time oscillations computed and plotted in one graph with the same base and scale Comparison between the short time oscillations of the two index series can then be made by studying the movements of the two curves

To compare long time changes the trend values of the two series should be plotted on the same base with the same scale and the directions of the resulting trend lines studied

EXERCISES

(1) Indicate briefly how you would analyze a series of monthly records extending over 50 years (M.A Alld 1942)

(2) (*a*) Explain fully what is meant by secular trend seasonal variations and cyclical fluctuations illustrating your answer

(*b*) Study the short time fluctuations of the following temperatures measured in degrees Fahrenheit —

Date 1941	Temperature	Date 1941	Temperature
Feb 1	40	Feb 11	78
2	50	12	[illegible]
3	44	13	60
4	70	14	64
5	[illegible]	15	62
6	44	" 16	68
7	30	17	86
8	40	18	96
9	56	19	94
10	68	20	78

(B Com Alld. 1942)

(3) Compare the long time changes and the short time oscillations of the following data

Year	Index No x	Index No y	Year	Index No x	Index No y
1900	80	102	1916	100	103
1	82	104	17	101	106
2	83	106	18	102	112
3	85	107	19	106	111

Year	Index No. x	Index No. y	Year	Index No. x	Index No. y
4	90	10	20	102	110
5	86	108	21	101	109
6	84	106	22	100	108
7	82	104	23	98	108
8	80	103	24	103	113
9	91	104	25	101	112
10	90	112	26	99	111
11	88	108	27	98	108
12	87	103	28	93	108
13	87	104	29	90	115
14	100	109	30	102	107
15	100	102	31	100	102

(4) (*a*) How would you distinguish the cyclical fluctuations from the trend and the seasonal fluctuations.

(*b*) The following table gives the value of the exports of merchandise from India during the years 1919-20 to 1923-24. Calculate the seasonal variations for each month during this period.

In Crores of Rupees

Months		1919-20	1920-21	1921-22	1922-23	1923-24
April	..	20	27	17	23	29
May	..	20	26	18	26	28
June	..	19	21	15	28	29
July	..	26	19	17	23	25
August	..	25	19	18	24	22
September	..	30	21	19	20	23
October	..	28	19	17	21	25
November	..	29	17	19	27	26
December	..	26	18	21	26	30
January	..	29	18	22	28	36
February	..	26	17	21	30	35
March	..	30	18	26	31	40

(M.A., Alld., 1937).

(5) The following table gives the Bank Clearings in the Bombay City for the years 1916 to 1940 in millions of rupees. Find the trend, and verify your result graphically.

1916	52.7	1929	94.6
1917	79.4	1930	83.0
1918	76.3	1931	110.6
1919	66.0	1932	159.6
1920	68.5	1933	177.4
1921	93.8	1934	178.6
1922	104.7	1935	235.8
1923	87.2	1936	243.2
1924	79.3	1937	194.4
1925	103.6	1938	217.9
1926	97.3	1939	214.0
1927	92.4	1940	256.7
1928	100.7		

(B Com Alld 1943)

(6) Classify the different types of fluctuations which occur in the analysis of time-series Illustrate your remark with the help of the following series

6099	6497	6898	7300	7699
6223	6621	7024	7421	7828
6351	6754	7152	7553	7949
6477	5878	7275	7075	8077

(M A, Cal, 1937)

(B Com. Nag 1944)

(7) Explain the use of moving averages in the analysis of time series Find out approximate moving average for the following series —

1901	506	1906	696	1911	1189	1916	898
1902	620	1907	1116	1912	818	1917	814
1903	1036	1908	738	1913	745	1918	920
1904	673	1909	663	1914	845	1919	1300
1905	588	1910	777	1915	1276	1920	961
						1921	906

(M.A Cal 1936)

(8) Write a note on the statistical analysis of time series in economic studies Illustrate your remarks with the help of the

following table, using in particular 3-year and 5-year moving averages :—

Year	Value	Year	Value	Year	Value	Year	Value
1901	507	1908	552	1915	583	1923	628
1902	522	1909	556	1916	581	1924	632
1903	524	1910	548	1917	599	1925	626
1904	521	1911	572	1918	6 2	1926	644
1905	538	1912	569	1919	597	1927	643
1906	541	1913	567	1920	612	1928	642
1907	537	1914	587	1921	616	1929	661
				1922	608	1930	659

(M.A., Cal., 1935).

(9) Draw a graph of the following time-series and study its trend :—

Year		Value	Year		Value
1910	..	496	1920	..	1442
11	..	615	21	..	1617
12	..	686	22	..	1678
13	..	835	23	..	1791
14	..	888	24	..	1916
15	..	1081	25	..	1883
16	..	1132	26	..	2064
17	..	1139	27	..	2278
18	..	1320	28	..	2368
19	..	1389	29	..	2345

(B. Com., Cal., 1937).

(10) Plot the following Index Numbers of wholesale price in U. S. A., and show the general trend of prices :—

Year			Index Number of Prices (1910—14=100)
1800	..	..	129
1810	..	..	131
1820	..	..	106
1830	..	..	91
1840	..	..	95

Year	Index Number of Prices (1910—14=100)
1850	84
1860	93
1870	135
1880	100
1890	82
1900	82
1910	103
1920	226
1930	126

(B Com Alld 1935)

(11) Business Cycles in the U S A and England arranged in chronological order (1796—1923) have had the following duration as measured to the nearest year —

U S A—
6 6 5 3 7 3 3 5 4 3 6 1 2 6 4 3 5 5
4 9 5 3 2 3 4 3 4 2 3 5 2 3

England—
4 6 4 3 5 4 6 4 2 6 10, 7 1 8 8 9 8 10,
7 6 5 2

Tabulate the above figures in classes of one year each and calculate the average duration of the business cycle in each country separately

(B Com Luck 1939)

(12) What is meant by trend? How would you statistically eliminate the influence of seasonal and cyclic factors on the long period movement of any series?

(B Com Bom 1936)

(13) The following are the Monthly Index Numbers of Commodity Group (Food and Tobacco) issued by the Economic Adviser to the Government of India

Prices for week ending 19th August 1939=100

Month	Index	Month	Index
1941		1942	
Oct.	127.4	July	155.8
Nov.	127.9	Aug.	158.9
Dec.	127.5	Sept.	161.0
1942		Oct.	167.2
Jan.	128.4	Nov.	172.4
Feb.	132.3	Dec.	178.5
March	130.5	1943	
Apr.	136.5	Jan.	190.8
May	144.7	Feb.	270.0
June	152.3		

Fit a straight line trend to the above data by the method of the Least Squares and exhibit the graph of the data with the trend.

(M. Com., Luck., 1944).

(14) Plot the following figures on a graph paper and study their trend. On a separate graph paper show their short-time oscillations with the trend removed.

Year		Value	Year		Value
1913-14	..	264	1924-25	..	305
-15	..	255	-26	..	303
-16	..	267	-27	..	306
-17	..	267	-28	..	297
-18	..	265	-29	..	292
-19	..	264	-30	..	304
-20	..	263	-31	..	310
-21	..	255	-32	..	317
-22	..	271	-33	..	331
-23	..	289	-34	..	344
-24	..	310	..	..	..

(15) Following are the total deposits of all exchange banks in India in crores of rupees Calculate five year and nine year moving averages and show them graphically

Year	Deposits	Year	Deposits
1915	34	1925	71
16	38	26	72
17	53	27	69
18	6	28	71
19	74	29	67
20	75	30	68
1	75	31	68
22	73	32	73
23	68	33	71
4	71	34	71
		35	76

(16) Fit straight line trends by the method of least squares to the following data spread over 10 consecutive years

(1)	(2)	(3)
1	20	110
2	19	125
3	18	115
4	17	135
5	16	150
6	15	165
7	14	155
8	13	175
9	12	180
10	11	200

(17) Calculate the seasonal variation Index from the data given in exercise (4) (*b*) page 395

(18) Distinguish between secular trends seasonal movements and cyclical fluctuations in dealing with a time series Explain carefully how secular trend should be measured

(M A Patna 1940)

(19) Explain fully what you understand by secular trend.

Assuming a ten-yearly cycle for the following series relating to Index Numbers of the Retail Price of Wheat in India (1873=100), give the trend values, and represent graphically the short-time fluctuations with the trend removed.

Year	Annual Average	Year	Annual Average	Year	Annual Average
1906	155	1914	200	1922	315
1907	168	1915	227	1923	356
1908	226	1916	193	1924	246
1909	203	1917	205	1925	294
1910	170	1918	270	1926	281
1911	153	1919	341	1927	267
1912	170	1920	310	1928	264
1913	177	1921	360	1929	262

(M. Com. Alld., 1944).

(20) Write a short essay on "Analysis of Time Series."

(B. Com., Hons , Andhra, 1943).

CHAPTER XVII

CORRELATION

Black cats cause bad luck while filled up pitchers good fortune—these are the beliefs held by some people But these beliefs are incapabl of being justified by mathematical theory It is therefore difficult to say if there really exists any relationship between black cats and bad luck and between filled up pitchers and good fortune though occasional coincidences may suggest such notions. On the other hand some people believe that devaluation of the rupee from 1s 6d rate to 1s 4d. would stimulate India s export trade or that a rise in the rate of interest would encourage savings These impressions do indicate some sort of relationship but they are mere guesses until they have been tested by the mathematical theory of drawing conclusions The theory by means of which quantitative connections between two sets of phenomena are determined is called the **Theory of Correlation**

Meaning of Correlation.

Correlation means a possible connection relationship or interdependence between two sets of phenomena If in each of them some factor is numerically measured and it is discovered that changes in the size of one factor run in sympathy with changes in the size of the other or to say the same thing large values of one go with large values of the other and small with small or *vice versa* the two factors exhibit some mutual dependence which is termed **correlation.** In other words if two quantities vary in sympathy so that a movement —an increase or decrease—in the one tends to be accompanied by a movement in the same or inverse direction in the other

and the greater the volume of change in the one the greater is the volume of change in the other, the quantities are said to be correlated.

In natural sciences correlation can be reduced to absolute mathematical terms. Heat always increases with light and an electric current is always associated with magnetic field. These instances suggest a high degree of correlation. But in social sciences it is seldom that any absolutely fixed mathematical relationship between two variables can be established. The law of demand, the law of diminishing returns, Gresham's law, to take a few illustrations, suggest correlation, but this correlation is not so perfect as that in the natural sciences. Therefore, in inexact sciences **we must take the fact of correlation established, if in a large number of cases two variables always tend to move in the same or opposite directions.**

Such phenomena are not uncommon in the social and economic sphere. We very often see that demand for a commodity generally falls with a rise in its price, that price level in a country generally rises with supply of money, that tall fathers generally have tall sons, that young husbands generally have young wives, that a taller man generally tends to be thinner. In all these cases correlation exists.

Positive and Negative Correlation.

Correlation may be positive or negative. If the two given variables steadily deviate in the same direction, correlation is direct or positive; but if they constantly deviate in the opposite directions, correlation is inverse or negative. That is, if an increase (or decrease) in the values of one variable is associated with an increase (or decrease) in the values of the other, correlation between them is positive. And, if an increase (or decrease) in the values of one variable is

associated with a decrease (or increase) in the values of the other correlation between them is negative One way of detecting the positive and negative character of correlation is to plot the two related variables on a graph paper that is draw correlation graphs and read the direction of the two curves. If they run parallel throughout (as they do in Figure 30 page 427) correlation is direct but if they run in opposite directions correlation is inverse If general level of prices rises with increase in the amount of money in circulation correlation between money in circulation and prices is positive. If with an increase in the production of sugar in India the imports of sugar have gone down the correlation between production and imports of sugar is inverse.

Degree of Correlation

Correlation exists in various degrees The radius of a circle bears a perfectly definite relationship with its area so that the area increases in a perfectly definite proportion with an increase in the radius Similarly the area of a square increases in a definite ratio with an increase in the length of its side These are the instances where correlation is perfect and positive Correlation will be perfectly negative if a fall of 10 per cent in the price of a commodity results in 10 per cent. rise in its demand. Similarly there may be instances where no correlation may exist If the height of a house is compared with that of a growing tree over a period of time it may be found that while the height of the house remained unchanged during the period that of the tree not only increased but also crossed that of the house Evidently the height of the house cannot be associated with that of the tree and therefore no correlation exists between them. Correlation may exist in a limited degree If demand for a commodity increases its price also increases but not necessarily in the same proportion This is a case of limited positive correlation If area under food

crops in a country increases, that under non-food crops may fall but not necessarily in the same proportion. This is an example of limited negative correlation.

Thus, correlation is **perfect positive** if an increase (or decrease) in one variable is always followed by a corresponding and proportional increase (or decrease) in the other related variable. It is **perfect negative** if an increase (or decrease) in one factor is followed by a corresponding and proportional decrease (or increase) in the other factor. There is **no correlation** at all if values in one variable cannot be associated with values in the other variable. In between perfect positive correlation and no correlation there may be **limited degrees of positive** correlation. Similarly, in between no correlation and perfect negative correlation there may be **limited degrees of negative** correlation.

Then, we may construct a scale which begins at the top with perfectly positive correlation, passes through limited degrees of positive correlation, reaches and crosses the entire absence of correlation, and passing through limited degrees of negative correlation ends at perfectly negative correlation. Such a scale is provided by Coefficient of Correlation.

Co-efficient of Correlation.

Coefficient of correlation is the numerical measure of the amount of correlation existing between two variables, subject and relative. That variable which is used as the standard is called the subject, and the variable which is compared with the subject or measured in terms of the subject is called the relative. Generally Karl Pearson's co-efficient of correlation is used. This coefficient varies between $+1$ and -1. When the coefficient reaches unity it is assumed to be perfect. Perfect positive correlation is indicated by $+1$, perfect negative by -1, no correlation

on complete independence by o and limited correlation by the intermediate values of the coefficient

Study of Correlation.

Correlation may be studied between (1) two *related* historical variables and (2) between any other two groups of *related* phenomena Correlation may for instance be studied between output of sugar in India and imports of it over a period of time to find whether with the increament in output in the country imports have fallen. It may be studied between supply of a commodity and its price over a period of time to find whether price falls with increase in supply These are examples of historical variables Correlation may be studied between the length and breadth of the leaves of a certain tree to find the relation between their length and breadth It may be studied between stature of fathers and stature of sons to find if tall fathers generally have tall sons. Both these are examples of non historical related phenomena. If however one produces figures to show that as the production of cane-sugar increased in India that of motor cars fell in the U S A. over a period of time or as the length of X leaf increased the breadth of Y leaf decreased these instances would not imply correlation unless there is reason to believe that production of cane sugar in India and of motor cars in the U S A are *related* in some way or the length of X leaf and breadth of Y leaf are groups of *related* phenomena

Karl Pearson's Co-efficient of Correlation.

To determine the degree of correlation between two related variables the co-efficient of correlation devised by Karl Pearson the great biologist and statistician is the most satisfactory This co-efficient is calculated by dividing the product of all the deviations of each pair of observations from their respective means by the product of the standard deviations of

the two variables and the number of items. Thus, if x_1, x_2, x_3, etc., be the deviations of the values of the first variable, the subject, from the arithmetic average, and y_1, y_2, y_3, etc., be the corresponding deviations of the values of the second variable, the relative, and the summation of the products of x_1 with y_1, of x_2 with y_2, of x_3 with y_3, and so on be represented by Σxy, and further the standard deviation of the subject be σ_1 and of the relative σ_2, and n be equal to the number of pairs of observations, then r, Karl Pearson's coefficient of correlation, will be

$$\frac{\Sigma xy}{n\ \sigma_1\sigma_2}$$

When Σxy is positive, correlation will be positive; when Σxy is negative, correlation will be negative. It is the numerator which largely regulates the size of the coefficient. If positive items in one series are associated with positive items in the other series, or if negative items in one series are associated with negative items in the other series, the coefficient of correlation is positive. This means that if items larger than the arithmetic average in the subject are associated with items larger than the arithmetic average in the relative, or items smaller than the mean in the subject are associated with items smaller than the mean in the relative, the correlation coefficient will be positive. If, however, positive items in one series are associated with negative items in the other or *vice versa*, the correlation co-efficient will be negative. When positive and negative deviations in the two series are indifferently associated, correlation will tend to zero, and will reach that limit when negative products of deviations will be equal to the positive products of deviations, *i.e.*, when Σxy will be zero.

Calculation of Pearson's Correlation Co-efficient: Direct Method.

Example 1. Required to calculate co-efficient of correlation between ages of husband and wife in a given community at a certain time, using the direct method.

Tabl 3 *Ca culation of Pearsonian Coefficient of Correlation between ages of husband and wife*

Subject X: Age of husband (years)	Deviation from average (25 years)	Square of deviation	Relative Y: Age of wife (years)	Deviation from average (18 years)	Square of deviation	Product of deviation of husband's age and of wife's age
m_1	x	x^2	m_2	y	y^2	xy
19	4	35	18	−4	16	+21
1	4	16	16	−2	4	+8
[illegible]	−3	9	15	−3	9	+9
3	−2	4	14	−4	16	+8
[illegible]	−2	4	17	−1	1	+2
21	−1	1	14	−4	16	+4
21	−1	1	17	−1	1	+1
[illegible]	0	0	18	0	0	0
26	+1	1	17	−1	1	−1
26	+1	1	20	+2	4	+2
27	+2	4	21	+3	9	+6
28	+3	9	20	+2	4	+6
29	+3	9	22	+4	16	+12
29	+4	16	22	+4	16	+16
30	+5	25	23	+5	25	+25
$\Sigma m_1 = 375$, $a = 25$		$\Sigma x^2 = 13$[illegible]	$\Sigma m_2 = 270$, $a_2 = 18$		$\Sigma y^2 = 13$[illegible]	$\Sigma xy = +122$

n or number of pairs of observations $= 15$

Standard deviation is determined by the formula[1], $\sqrt{\frac{\Sigma d^2}{n}}$. In table 53 the d's in the X series are called x's, those in Y series, y's. Accordingly the formula for computing the standard deviation of the X series is $\sqrt{\frac{\Sigma x^2}{n}}$ for that of the Y series $\sqrt{\frac{\Sigma y^2}{n}}$

[1] See Table 23 Chapter XI page 193 for computing standard deviation the Direct Method.

$$\therefore\ \sigma_1=\sqrt{\frac{\Sigma x^2}{n}}=\sqrt{\frac{136}{15}}=3.01 \text{ years.}$$

$$\text{and } \sigma_2=\sqrt{\frac{\Sigma y^2}{n}}=\sqrt{\frac{138}{15}}=3.03 \text{ years.}$$

$$r=\frac{\Sigma xy}{n\sigma_1\ \sigma_2}$$

$$=\frac{+122}{15\times3.01\times3.03}=+.89.$$

.89 indicates a very high degree of positive correlation, implying that the age of wife increases with that of husband.

Calculation of Pearson's Co-efficient: Short-cut Method.

In the above example the averages of the ages of husbands and wives were whole numbers. Therefore, the calculation of the deviations of ages, the mean, their squaring up, and their multiplication did not involve any trouble. If, however, the averages contain a fraction, these calculations would involve much labour, to do away with which the short-cut method may be used. In using it, any whole number may be assumed as the average, deviations from it calculated, and squared, and the standard deviations computed according to the short-cut method of computing the standard deviation. The deviations of the two series are multiplied and summated. The resulting xy should later be corrected by subtracting from it the product of the differences between the true means and the assumed means of the two series. Thus, if p be the true average of the products, *i e*, true or corrected value of Σxy, then

$$p=\frac{\Sigma xy}{n}-\left[\ (a_1-x_1)\ (a_2-x_2)\ \right]$$

where, a_1 stands for the true average and x_1 for the assumed average of the first series, and a_2 stands for the true average and x_2 for assumed average of the second series, and Σxy is the summation of the products of deviations from assumed means.

Then the co-efficient of correlation or, $r = \dfrac{p}{\sigma_1 \, \sigma_2}$

The above two processes may also be combined into one formula so that without changing what the symbols stand for

$$r = \frac{\Sigma xy - n\,[(a_1 - x_1)\ (a_2 - x_2)]}{n\ \sigma_1 \sigma_2}$$

Example 2 Required to calculate the co-efficient of correlation between birth rate and death rate of a few countries of the world during 1931 using the short-cut method

Table 54 *Calculation of the Pearsonian Co-efficient of Correlation between birth rate and death rate for a few countries of the world for 1931*

Country	Birth rate	Deviation from assumed average (26)	Square of deviation	Death rate	Deviation from assumed average (15)	Square of deviation	Product of deviations of birth rate and death rate
	m	x	x^2	m_2	y	y^2	xy
Egypt	44	+18	324	27	+12	144	+216
Canada	24	− 2	4	11	− 4	16	+ 8
U S A	19	− 7	49	12	− 3	9	+ 21
India	33	+ 7	49	24	+ 9	81	+ 63
Japan	32	+ 6	36	19	+ 4	16	+ 24
Germany	16	−10	100	11	− 4	16	+ 40
France	18	− 8	64	16	+ 1	1	− 8
I F State	20	− 6	36	14	− 1	1	+ 6
U K.	16	−10	100	12	− 3	9	+ 30
U S S R	40	+14	196	18	+ 3	9	+ 42
Australia	20	− 6	36	9	− 6	36	+ 36
Newzealand	18	− 8	64	8	− 7	49	+ 56
Palestine	53	+27	729	23	+ 8	64	+216
Sweden	15	−11	121	12	− 3	9	+ 33
Norway	17	− 9	81	11	− 4	16	+ 36
$n = 15$	$\Sigma m = 385$ $a_1 = 25.67$	-67 +72 +5	$\Sigma x^2 = 1959$	$\Sigma m_2 = 227$ $a_2 = 15.13$	-25 23 (9)	$\Sigma y^2 = 476$	$\Sigma xy = +817$

n_2 or number of pairs of observations $=15$.

a_1 or True arithmetic average for the first series

$$=\frac{\Sigma m_1}{n}=\frac{385}{15}=25.67$$

Let x_1, or Assumed average for the first series $=26$

a_2 or True arithmetic average for the second series

$$=\frac{\Sigma m_2}{n}=\frac{227}{15}=15.13$$

Let x_2 or Assumed average for the second series $=15$ Standard deviation, using the short-cut method, is determined by the formula, $\sqrt{\frac{\Sigma d^2{}_x - n(a-x)^2}{n}}$ in table 54, d_x's in the first series are called x's; those in the second series y's. Accordingly, the formula for computing standard deviation of *the first series is* $\sqrt{\frac{\Sigma x^2 - n\ (a_1-x_1)^2}{n}}$ *and for that of the* second series, $\sqrt{\frac{\Sigma y^2 - n(a_2-x_2)^2}{n}}$

$$\sigma_1=\sqrt{\frac{1989-15\ (25.67-26)^2}{15}}=11.5$$

$$\text{and}\ \sigma_2=\sqrt{\frac{476-15\ (15.13-15)^2}{15}}=5.612$$

$$r=\frac{\Sigma xy - n\ [(a_1-x_1)\ (a_2-x_2)]}{n\ \sigma_1\ \sigma_2}$$

$$=\frac{+819-15\ (-.33\times.13)}{15\times11.5\times5.612}$$

$$=+.848.$$

+.848 denotes a very high degree of positive correlation between birth-rate and death-rate of the given countries of the world.

[1] See Table 25, Chapter XI, page 201, for computing standard deviation by the Short-cut Method.

Co efficient of Correlation for Long-Time Changes

In the above two examples the variations in the items relate to a specific time. Correlation may also be studied for historical data, that is, data stretched over a period of time. Historical data may relate to (1) Long time changes and (11) Short time oscillations In computing the co-officient of correlation for long time changes the method used in example 1, or if need be, used in example 2 shall be followed throughout the items and deviations from the mean for the same date being paired together In computing the co-efficient of correlation for short time oscillations this method will be modified

Pearson's Modified Co-efficient for use with Short-Time Oscillations.

It is possible that the short time changes in two variables may be in opposite directions while the long term changes may be in the same direction Then, if co-efficient of correlation of such variables is computed by the method used in the foregoing two examples, a large positive co-efficient would result which would not take any account of the opposite direction of the short time oscillations Correlation co-efficient computed from actual items would consequently be misleading We *should therefore, be concerned with short time oscillations only* and rid our data of the long time variations* To do it we should discover the trend and eliminate it by computing the deviations of original items from the trend These deviations should be multiplied together to yield Σxy And these deviations again should be squared up to compute standard deviations Thus the modification made in the original formula is that deviations of the items are taken from the trend instead of from the arithmetic average Example 3 demonstrates the working of this method

* See Chapter XVI for 'Elimination of long time var ations.'

Example 3. Required to compute the co-efficient of correlation of the short-time oscillations for indices of supply and price of a certain commodity.

Table 55. *Computation of Co-efficient of Correlation of Short-time Oscillations between Supply and Price.*

Year	Supply X: Index of Supply	Supply X: 5-yearly moving average of Indices	Supply X: Deviation from moving average x	Supply X: Square of deviation x^2	Price Y: Index of Price	Price Y: 5-yearly moving average of Indices	Price Y: Deviation from moving average y	Price Y: Square of deviation y^2	Product of deviations of Supply and of Price xy
1920	91				117				
1921	98				97				
1922	95	94	+1	1	102	106	−4	16	−4
1923	92	95	−3	9	108	102	+6	36	−18
1924	93	96	−3	9	105	98	+7	49	−21
1925	96	98	−2	4	96	91	+5	25	−10
1926	102	100	+2	4	77	85	−8	64	−16
1927	107	101	+6	36	68	82	−14	196	−84
1928	104	102	+2	4	77	81	−4	16	−8
1929	98	103	−5	25	93	82	+11	121	−55
1930	100	105	−5	25	89	84	+5	25	−25
1931	108	107	+1	1	83	85	−2	4	−2
1932	116	110	+6	36	78	85	−7	49	−42
1933	114				84				
1934	111				93				
		$n=11$		$\Sigma x^2=154$		$n=11$		$\Sigma y=601$	$\Sigma xy=285$

[Five-yearly cycle has been assumed in the above series and decimals have been ignored in computing the moving average. Greater precision could be achieved by carrying out the decimals.]

n or number of pairs of observations $=11$, since only the years 1922 to 1932 can be used in computing the co-efficient.

$$\sigma_1=\sqrt{\frac{\Sigma x^2}{n}}=\sqrt{\frac{154}{11}}=\sqrt{14}=3.742$$

$$\sigma_2 = \sqrt{\frac{\Sigma y^2}{n}} = \sqrt{\frac{601}{11}} = \sqrt{54.6} = 7.389$$

$$r = \frac{\Sigma xy}{n\ \sigma_1\ \sigma_2} = \frac{-85}{11 \times 3.742 \times 7.389} = -.937$$

—.937 denotes a very high degree of inverse correlation between supply and price for short period indicating that as supply increases price falls and *vice versa*

Calculation of Correlation Co-efficient in Grouped Series.

In the foregoing three examples the given series relate to quantities and dual observations Correlation of grouped series can also be similarly studied We may measure an adequate number of pairs of values for each member and find what values are associated together and how often the same values are repeated. When this is done we can group our data into a table of double entry or contingency table Suppose we find that in two class tests—one in Economics and the other in Geography—at which 60 boys were examined the following were the results —

Table 56 *Frequency distribution of Marks in Economics & Geography*

X Economics (Max. Marks 50)		Y Geography (Max Marks 50)	
Marks obtained	Number of boys	Marks obtained	Number of boys
5—15	5	0—10	2
15—25	18	10—20	15
25—35	27	20—30	20
35—45	10	30—40	15
		40—50	8
	60		60

If we desire to study the relationship between the knowledge of Economics and that of Geography with the help of the above two series, we would need some more information: we should know what values of the two series are associated together and how frequently the same values are repeated. Suppose we find that one boy who got marks varying between 5—15 in Economics also got marks varying between 0—10 in Geography, that three boys who got marks varying between 5—15 in Economics also got marks varying between 10—20 in Geography, and so on, we can prepare a table of double entry as follows:—

Table 57. *Correlation Table for marks in Economics and Geography.*

Y Marks in Geography (max. marks 50)	X Marks in Economics (max. marks. 50)				Total
	5—15	15—25	25—35	35—45	fy
0—10	1	1			2
10—20	3	6	5	1	15
20—30	1	8	9	2	20
30—40		3	9	3	15
40—50			4	4	8
Total fx	5	18	27	10	60

Table 57 shows the grouped frequency distribution of two variables. This distribution may be termed as Bivariate Fre-

quency Distribution and the table as Contingency Table But if we are particularly interested in the relationship between the two variables this table of double entry may be designated as **Correlation Table.**

Example 4 Required to compute correlation co-efficient from the data given in table 57

[To compute the co-efficient the formula used in example 1 where deviations were calculated from the true mean, or the formula used in exercise 2 where deviations were calculated from the assumed mean may be used in this example too The later procedure saves much labour and, therefore, it will be adopted in the given case.]

In tables 58 and 59 below we calculate the standard deviations of the X and the Y series relating to marks in Economics and Geography, respectively Let the assumed averages, x_1 and x_2 for the X and the Y series be respectively 30 and 25 marks

Table 58 *Calculation of Standard Deviation of X series*

Marks group	Mid value m	Frequency f	Product of mid value & frequency mf	Deviation from assumed average (30) d_x	Square of deviation d^2_x	Product of frequency & square of deviation fd^2_x
(1)	(2)	(3)	(4)	(5)	(6)	(7)
5—15	10	5	50	—20	400	2000
15—25	20	18	360	—10	100	1800
25—35	30	27	810	0	0	0
35—45	40	10	400	+10	100	1000
		$n=60$	$\Sigma mf=1620$			$\Sigma fd^2_x=4800$

x_1 or Assumed Average$=30$ marks.

a_1 or True Average$=\dfrac{\Sigma m}{n}=\dfrac{1620}{60}=27$ marks

$$\sigma_1=\sqrt{\frac{\Sigma d^2{}_x-n\,(a_1-x_1)^2}{n}}$$

$$=\sqrt{\frac{4800-60\,(27-30)^2}{60}}$$

$$=\sqrt{71}=8.425 \text{ marks.}$$

Table 59. *Calculation of Standard Deviation of Y series.*

Marks group	Mid-value	Fre-quency	Product of mid-value & frequency	Deviation from assumed mean (25)	Square of deviation	Product of frequency & square of deviation
	m	f	mf	d_y	$d^2{}_y$	$fd^2{}_y$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0—10	5	2	10	—20	400	800
10—20	15	15	225	—10	10	1500
20—30	25	20	500	0	0	0
30—40	35	15	525	+10	100	1500
40—50	45	8	360	+20	400	3200
		$n=60$	$\Sigma m=1620$			$\Sigma d_y=7000$

x_2 or Assumed Average$=25$ marks.

a_2 or True Average$=\dfrac{\Sigma m}{n}=\dfrac{1620}{60}=27$ marks.

$$\sigma = \sqrt{\frac{\Sigma d^2 - n\,(a - x)^2}{n}}$$

$$= \sqrt{\frac{7100 - 60\,(27 - \cdot 5)^2}{60}}$$

$$= \sqrt{112\,66} = 10\,614 \text{ marks}$$

Now the value of Σxy remains to be determined Table 60 shows the method of determining it d_x, deviations from the assumed mean in X series shown in column (5) table 58 and d_y deviations from the assumed mean in Y series shown in column (5) table 59 are taken to table 60 Related pairs of deviations are first multiplied and the product put down in the left hand top corner of their respective cells Thus —20 d_x is multiplied with —20 d_y, and the product placed in the left corner of the cell formed by the 1st row and the 1st column —20 d_x and —10 d_y are multiplied and the product placed in the left corner of the cell formed by the 2nd row and the 1st column —10 d_x and +10 d_y are multiplied and the product—100 placed in the left corner of the cell formed by 4th row and 2nd column and so on

These products of d_x and d_y are multiplied by their respective frequencies placed in the centre of their respective cells The final products are placed in the right hand bottom corner of their respective cells These final products when summated give the Σxy This summation refers to the assumed averages and will therefore, as in example 7 be corrected by subtracting from it n times the product of the difference between true and assumed means of the two series

Table 60. *Calculation of Summation of Products of Deviations,* (Σxy)

Column No.			1	2	3	4	
Row No.	Marks → X ↓ Y	d_x → d_y ↓	5—15 —20	15—25 —10	25—35 0	35—45 +10	Product of d_x and d_y and frequency fxy
1	0—10	—20	400 / 1 / 400	200 / 1 / 200	—	—	690
2	10—20	—10	200 / 3 / 600	100 / 6 / 600	0 / 5 / 0	—100 / 1 / —100	1100
3	20—30	0	0 / 1 / 0	0 / 8	0 / 9 / 0	0 / 2 / 0	0
4	30—40	+10	—	—100 / 3 / —300	0 / 9 / 0	100 / 3 / 300	0
5	40—50	+20	—	—	0 / 4 / 0	200 / 4 / 800	800
	Product of d_x and d_y and frequency. fxy		1000	500	0	1000	Σy=2500

$$r = \frac{\Sigma xy - n\,[(a_1 - x_1)\ (a_2 - x_2)]}{n\sigma_1\ \sigma_2}$$

$$= \frac{2500 - 60\,[(27-30)\ (27-25)]}{60 \times 10.614 \times 8.486}$$

$$= \frac{2860}{5403.587}$$

$$= +.53$$

+ 53 ind cates a moderately h gh degree of positive correla tion between Economics and Geography

In the above example it was assumed that the values of the various frequencies in the X and the Y series were equal to the mid values of their class intervals Accordingly the deviations standard deviations and products of deviations had reference to the mid values of marks groups No doubt a particular class interval includes all values between its class lim ts but the assumption we have made does not generally create a large d fference in the result and is usually adopted.

In examples 1 2 and 4 correlation is positive but no perfect A simple example of perfect positive correlation is the following —

Number of persons	1,	2	3	4	5	6	7	8
Number of eyes	2	4	6	8	10	12	14	16

Correlation co-efficient in the above series will be + 1

Assumptions of Pearsonian Correlation.

Karl Pearson s co-efficient of correlation is based on two assumptions

(1) *In each of the series correlated a large variety of independent causes are operating so as to produce normal distribution* Such causes for example are variations in climate nourishment phys cal training environment The series resulting from the effect of such independent contributory causation would show normal d stribution Such causes were for instance operating in the determination of ages of husbands and wives in example 1

(2) *The forces so operat ng are related in a causal way* If the forces are independent of each other there would be no correlation If the height of a house remained unaltered while that of a growing ch ld increased there would be no correlation bet veen them s nce the causes affect ng one variable would

not be found to affect the other, that is, the sizes in one could not be said to be associated with the sizes in the other.

Characteristics of Pearsonian Co-efficient.

Karl Pearson's co-efficient of correlation is zero when independence between two variables is complete and is unity when there is perfect correlation, *i.e.*, when connection between the variables is rigid. It always varies between $+1$ and -1 and is a sensitive measure of the amount of correlation. It is based on all the observations of the given variables and is independent of the units in which the variables are measured.

Probable Error of the Co-efficient.

Probable error is a measure which when added to or subtracted from a most probable measurement gives the limits within which it is probable that an item of the same nature, if selected *at random*, will fall.

Co-efficient of correlation also has a probable error. It is that amount which when added to and subtracted from the average correlation co-efficient gives amounts within which the probability is that a co-efficient of correlation from series selected *at random* from the same universe will fall.

The formula for the probable error of Karl Pearson's co-efficient of correlation is

$$.6745 \frac{1-r^2}{\sqrt{n}}$$

where, r is the co-efficient of correlation and n the number of items paired.

The probable error of the co-efficient of correlation, $+.89$ between the ages of husband and wife computed in example 1 will be

$$.6745 \frac{1-(.89)^2}{\sqrt{15}} = .036.$$

The co-efficient of correlation for the example under consideration should, therefore, be written as

$$r = +.89 \pm .036$$

That is the co-efficient of correlation lies between .926 and .854

Interpretation of Co-efficient of Correlation

As already pointed out the co-efficient of correlation for a given pair of variables must lie between +1 (perfect direct correlation) and −1 (perfect inverse correlation). When the co-efficient is zero it indicates complete absense of correlation. So when the co-efficient is 0.9 or over the degree of correlation is obviously high.

It may be asked whether correlation between a given pair of variables as indicated by the co-efficient is significant. This can be answered in the light of the following four empirical rules based on the use of probable error of the co-efficient.

(1) If r is less than the probable here is no evidence of correlation.

(2) If r is more than six times the probable error correlation is significant that is its existence is a practical certainty.

(3) If the probable error is relatively small correlation should not be considered at all marked when r is less thon 0.30.

(4) If the probable error is relatively small correlation is decidedly existing when r is above 0.50.

In example 1 r is nearly 25 times the probable error. Correlation is therefore significant. We can now say that the co-efficient of correlation in the example actually lies between .926 and .854 and that another co-efficient computed from series chosen *at random* from the universe from which the given series was selected would fall within this range.

It may be noted that the probable error at times leads to wrong results unless r is small and n is large. In order that

the formulae for co-efficient of correlation and probable error may yield satisfactory results, n should be considerably large.

The above four rules must be kept in view while interpreting the correlation co-efficient. When a correlation co-efficient, it may be added, is found to be significant, it should not be implied to mean more than what it does. For instance, in example 1, correlation between ages of husbands and wives is strong positive. It simply shows a connection between the two age series and does not necessarily mean that *every* young husband has young wife. **Correlation. is true on the average.** A particular old man may have a young wife or two wives, one young and the other old. Again, if supply and price in example 3 are negatively correlated, it does not mean that increase in supply is the *only* cause of fall in price. There may be several other causes too leading to this particular 'effect'. Similarly, if marks in Geography and Economics are positively correlated it does not imply that the two subjects are necessarily related as cause and effect. Knowledge of one subject may be helpful in the other, but the correlation may also be due to some third factor, *e.g*, adequate teaching in both the subjects. So a **direct cause and effect relationship is not always and in all cases established by the fact that two series are correlated.**

Co-efficient of Concurrent Deviations.

So far we were concerned with only one method of measuring correlation which may be termed as the "Sum Product" method, since the measure is dependent on the sum of the products of the deviations. If, however, a measure of association in the *direction of change* alone is desired, the method of concurrent deviations may be used.

In example 3, we have used the modified method of measuring co-efficient of correlation for short-time oscillations.

Correlation by Graphic Method

While discussing positive and negative correlation it was pointed out that one way of detecting the negative or positive character of correlation is to draw correlation graphs and read the direction of the curves. This method is illustrated in Figure 30 below in which monthly figures relating to volume and value of exports of rice by sea from India given in table 62 are plotted.

Table 62. *Foreign Seaborne Trade—Exports (Value and Volume) of Rice (not in husk) in 1941-42.**

Month	Volume	Value
	Tons (000)	Rs (00 000)
April	[illegible]	32
May	29	43
June	22	38
July	19	29
August	27	44
September	43	69
October	24	40
November	18	30
December	20	31
January	23	37
February	32	53
March	29	43
Average	25.4	40.3

In drawing correlation graphs the choice of scales and base line should be so made that if lines representing averages of the two series are drawn parallel to the base they would be as close to each other as possible. There is no objection to taking a false base line if it is required for bringing the two average lines nearer each other. By drawing the curves on

* Compiled from the *Monthly Survey of Business Conditions in India* August 1942

such base line and scale their fluctuations would be thrown into proper relief.

Curves representing volume and value of rice exported from India

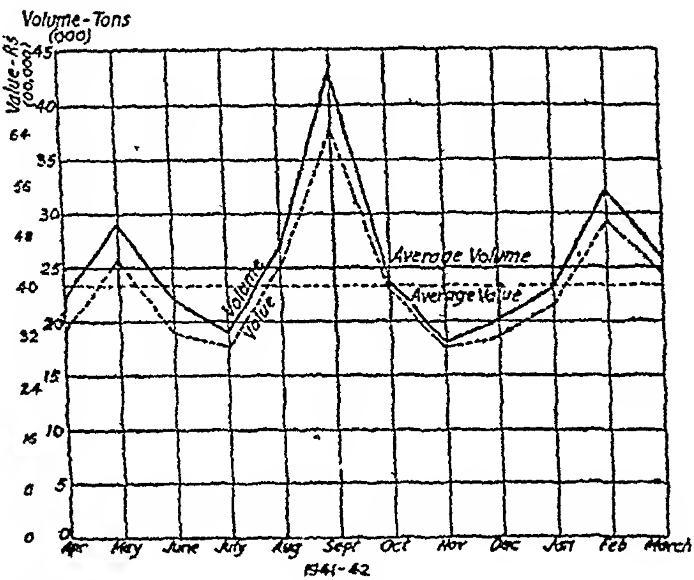


Fig. 30

In Figure 30 the lines showing average volume and average value are close to each other. They may have been brought still closer or made to overlap each other. But, thereby the two curves, the nature of their series being such, may also have overlapped each other, at least for a large part, and their graphic display would have been spoilt. The curves, as they are drawn, serve their purpose alright. They run remarkably parallel to each other throughout their upward and downward journey. They, therefore, indicate positive correlation between volume and value of exports of rice during the several months of 1941-42. Similarly, with another correlation graph we may have found

that both the curves steadily ran in opposite direction, yielding negative correlation between the series

But correlation graphs are not capable of doing anything more than suggesting the fact of a possible relationship between two variables We can certainly note from them whether fluctuations agree throughout their courses, whether both of them rise and fall together, whether maxima and minima occur at the same dates and so on but we can neither establish any causal relationship between the two variables nor obtain the exact degree of correlation through them They only tell us whether the two variables are positively or negatively correlated

It may be observed that in investigating causal relations ratios help more than quantities If two variables are really related to each other, the proportional increase or decrease in one may vary direcly with the proportional increase or decrease in the other Consequently, resemblance between two curves may be brought out distinctly if the logarithmic scale is substituted for the natural scale The same can also be done by reducing the two variables to index numbers and plotting them on a graph with common base line and common vertical scale Preference may be given to logarithmic scale over the natural scale for the additional reason that wrong and deceptive conclusions might be drawn if scale are shrewdly manipulated and base lines not inserted correctly while using the natural scale

If two given series show fluctuations with time and we are interested in the correlation for long time and for short time changes a different method of comparison will be followed

Graphic Correlation of Time Changes.

In example 3 we have used a modification of Pearson's method for computing the correlation co-efficient The reason for the modification is that if we desire to discover the

relationship of short-time oscillations we should rid our data of the long-time variations. If the annual index numbers of supply and price in table 55, page 413, are plotted on a graph and their correlation is studied, it would be found that it is negative, the wo curves moving in opposite directions. If long-period changes are compared by plotting the 5-yearly averages given in columns 3 and 7 it will be seen that as supply shows a rising trend from 1920 onwards the price shows a downward trend from 1920 upto 1928; but, after 1928 even with an upward trend of supply the trend of the price is also upward, which fact may be due to increase in population, prosperity or change in-demand. A comparison of the long-time changes after 1928 would, therefore, suggest a positive correlation, while that of those before 1928, a negative correlation. But when the long-time variations are eliminated and we are left with short-time oscillations, as given in columns 4 and 8, we may plot these deviations on a graph (as we did in Figure 29) and observe a marked relationship between the two curves, now completely unobscured by long-time changes. We would see that when one curve rises the other falls and *vice versa* The fact of negative correlation would thus be made clear.

It follows, therefore, that to study the correlation for long-time changes in a time series we should plot the moving averages of the two series and compare their directions, while to study the correlation for short-time oscillations we should plot the deviations from the trend and observe their movement.[5]

Lag and its Use.

Not infrequently it is found that while two given historigrams exhibit similar movements thereby indicating the existence of correlation, the crests and troughs of the waves in the graphs

---

[5] See in this connection 'Comparison of time changes in two historigrams,' Chapter XVI, page 393.

do not coincide one curve lagging behind the other This may be due to the fact that the effects follow the 'causes' only after an interval of time has elapsed between them For instance, increament in the supply of money causes a rise in the general level of prices which in its turn leads to a rise in the cost of living but it takes time for rise in the cost of living to result from increament in the supply of money

When lag is known to exist it will be well to allow for it by lagging the effect curve so that the curve may be made comparable with the cause curve. An easy way of determining the period of lag is to plot the two series on one graph—i e draw correlation graph—and read the time-distance between the peaks of the two curves Suppose cost of living index shows a rise three months after the rise in the index of supply of money then the period of lag is three months To allow for it the cost of living curve should be lagged by three months Thus in computing the co-efficient of correlation the index of money for January should be paired with the index of cost of living for April index of money for February with index of cost of living for May and so on The underlying principle is that the dates coupled together should be not identical but so different that the resulting correlation will be the highest

EXERCISES

(1) Discuss fully what is meant by the co-efficient of correlation and how it is measured and interpreted

(B Com Alld 1942)

(2) Define correlation coefficient What inferences can you draw from the values +1 0 and —1 of this co-efficient?

(3) What is correlation? Explain how you will use the following methods in determining correlation:—

(i) Graph, (ii) Correlation table, (iii) Karl Pearson's Co-efficient of Correlation.

(B. Com., Agra, 1940)

(4) Find graphically if the volume and value of imports of liquor data for which are given in exercise 14, Chapter XV, page 376, are related to each other.

(5) Find graphically if the volume and value of imports of capital outlay and gross earnings from the data given in exercise 12, Chapter XV, page 375.

(6) Find the correlation between exports and imports for 1920-21 data for which are given in exercise 7, Chapter XV, page 373.

(7) Compute the co-efficient of correlation of the short-time oscillations from the data relating to index numbers of X and Y (for the 1st 16 years only) given in exercise 3, Chapter XVI, page 394. Assume 5-yearly cycle and ignore decimals.

(8) Find the Pearsonian co-efficient of correlation between Index Numbers of Prices of Cotton Shares and those of Coal Shares from the data given in Exercise 24, Chapter XI, page 219.

(9) $\sigma_1=4.5$ and $a_2=3.6$ are the standard deviations of two groups $x_1, x_2, x_3, \ldots x_n$ and $y_1, y_2, y_3 \ldots y_n$ and $\Sigma xy=4800$, $n=1000$.

Calculate the co-efficient of correlation between the above two groups and interpret it. Also give the probable error of the co-efficient.

(10) What is meant by the probable error of co-efficient of correlation? Why and how is it measured?

(11) Calculate the co-efficient of correlation between the total receipts and the passengers given in exercise 27, Chapter XI, page 220.

(B. Com., Alld., 1932).

(1[illegible] The index numbers of prices of all com[illegible]od[illegible]t[illegible]s [illegible] Bom[illegible]b[illegible] and in Calcutta were as under —

Month	Index number of Commodity prices in Calcutta	Index number of Commodity prices in Bomb[illegible]
May 194[illegible]	169	[illegible]04
June 194[illegible]	18[illegible]	[illegible]
July 1942	18[illegible]	[illegible]
August 194[illegible]	19[illegible]	[illegible]8
S[illegible]p[illegible]ember 194[illegible]	198	229
October 1942	[illegible]09	[illegible]33
November 194[illegible]	[illegible]	249
December 194[illegible]	[illegible]38	[illegible]66
January 1943	[illegible]50	[illegible]55
February 1943	[illegible]53	[illegible]5[illegible]

Do you think prices in Bombay and in Calcutta are correlated

(M [illegible] Agra 1944)

(13) The following table gives the value of exports of raw cotton from India and the value of the imports of manufactured cotton goods into India during the years 1913 14 to 1931 3[illegible] —

(In Crores of Rupees)

Year	Exports of Raw Cotton	Imports of manufactured Cotton Goods
1913 14	4[illegible]	56
191[illegible] 18	44	49
1919-[illegible]0	58	53
1921 [illegible]	55	58
19[illegible]3 24	89	6[illegible]
19[illegible]9-30	98	76
1931 32	66	58

Calculate the co-efficient of correlation between the value of the exports of raw cotton and the value of the imports of cotton manufactured goods

(M.A. Cal. 193[illegible])

(B Com. Nag 1945)

(14) Calculate the co-efficient of correlation from the following data:

Amount of cheques cleared in Calcutta and Bombay Clearing Houses.

		Crores of Rs.				Crores of Rs.	
Year		Calcutta	Bombay	Year		Calcutta	Bombay
1925	..	1018	519	1933	..	824	646
26	..	959	421	34	..	864	683
27	..	1024	398	35	..	939	750
28	..	1088	543	36	..	899	721
29	..	998	800	37	..	993	837
30	..	893	712	..		..	..
31	..	756	640	..		..	..
32	..	747	646	..		..	..

(15) The following table gives the wholesale price index numbers for Calcutta and Karachi for the period 1927-1941:—

Year	Calcutta Index Numbers (Base: July, 1914)	Karachi Index Numbers (Base: July, 1914)
1927	148	137
1928	145	137
1929	141	133
1930	116	108
1931	96	95
1932	91	99
1933	87	97
1934	89	96
1935	91	99
1936	91	102
1937	102	108
1938	95	104
1939	108	108
1940	120	116
1941	139	120

(*a*) Calculate the co-efficient of correlation between the above two series, and state what it indicates.

(*b*) Point out whether the Calcutta indices are more variable than the Karachi ones.

(B. Com., Alld., 1944).

(16) Calculate the co efficient of correlation between the cost of living and the weekly wage rates from the following data —

Date	Cost of Living Index	Index of Weekly Wage Rates
1920	151	155
1921	110	120
1922	102	99
1923	101	98
1924	103	101
1925	100	101
1926	100	102
1927	96	100
1928	95	99
1929	95	99
1930	87	98
1931	84	96
1932	81	94

(M.A., Alld, 1937)

(17) The following table gives the number of students having different heights and weights

Height in Inches	Weight in pounds					Total
	80—90	90—100	100—110	110—120	120—130	
50—55	1	3	—	5	2	11
55—60	2	4	10	7	4	27
60—65	1	5	12	10	7	35
65—70	—	3	8	6	3	20
Total	4	15	30	28	16	100

Do you find any relation between height and weight?

(B Com, Alld 1940)

✓(18) The following table gives the index numbers of industrial production of Great Britain and the number of registered unemployed persons in the same country during the years 1924-1931 :—

Year	Industrial Production (Index Number)	Number of registered Unemployed (Hundred thousand)
1924	100	11.3
1925	102	12.4
1926	104	14.0
1927	107	11.1
1928	105	12.3
1929	112	12.2
1930	103	19.1
1931	94	26.4

Calculate co-efficient of correlation between production and the number of unemployed.

(B. Com., Luck., 1944).

✓(19) Find the correlation co-efficient between heights of father and son from the following data :—

Height of father in inches ..	65	66	67	67	68	69	71	73
Height of son in inches ..	67	68	64	68	72	70	69	70

(M.A., Alld., 1940).

(20) Find the co-efficient of correlation between marks obtained by candidates at an examination in two subjects A and B from the following data —

Subject A Max. 50	Subject B—Maximum 50					Total
	11—15	16—20	21—25	26—30	31—35	
1—5					1	1
6—10	1	1	8	7	1	18
11—15	1	2	4	14	4	25
16—20			7	13	6	26
21—25			2	4	1	7
26—30			1			1
31—35				1		1
Total	2	3	23	39	12	9

(B Com Bom. 1936)

(21) The following table gives the frequency, according to age-groups of marks obtained by 65 students in an intelligence test —

Test Marks	Age in years				
	19	20	21	22	Total
200—250	4	4	2	1	11
250—300	3	5	4	2	14
300—350	2	6	8	5	21
350—400	1	4	6	8	19
Total	10	19	20	16	65

Is there any relation between age and intelligence?

✓ (22) What are the assumptions upon which the Pearsonian co-efficient of correlation is based? How does the positive correlation differ from the negative? Compute the co-efficient of correlation of the short-time oscillations from the following data:—

Year		Supply	Price
1921	..	80	146
1922	..	82	140
1923	..	86	130
1924	..	91	117
1925	..	83	133
1926	..	85	127
1927	..	89	115
1928	..	96	95
1929	..	93	100

(Assume a three-year cycle, and ignore decimals).

(M. Com., Alld., 1943).

✓ (23) From the following table, find out how far the fluctuations in prices correspond to the amount of money in circulation in India:—

Year		Rupees and Notes in Circulation in crores	Index Number of Prices (1873=100)
1912	..	248	137
1913	..	256	143
1914	..	248	147
1915	..	266	152
1916	..	297	184
1917	..	338	196
1918	..	407	225
1919	..	463	276
1920	..	411	281
1921	..	393	260

(B. Com., Agra, 1937).

(24) Find the co-efficient of correlation from the following table

$x=$ / $y=$	5	10	15	20	25	30	Total
10		1	1	2	8	12	24
15	1	2	5	9	80	11	108
20	2	15	42	98	36	8	201
25	5	20	51	37	10	2	125
30	8	16	8	5	4	1	42
Total	16	54	107	151	138	34	500

(M.A, Cal, 1937)

(25) The following table shows the distribution of marks. Calculate the co-efficient of correlation and its probable error —

Marks in Geography

Marks in Mathematics — Range of Marks	0—20	20—40	40—60	60—80	Total
0—20	32	88	15	—	135
20—40	45	436	200	4	685
40—60	16	500	308	25	939
60—80	—	105	532	40	677
80—100	—	8	40	16	64
Total	93	1 137	1 185	85	2 500

(M.A, Cal 1935)

(26) What is meant by correlation? Give the general rules for interpreting its co-efficient

Calculate the co-efficient of correlation between the ages of 100 husbands and wives from the following data :—

Age of Husbands in years.	Age of wives in years.					Total
	10—20	20—30	30—40	40—50	50—60	
15—25	6	3	..	..	..	9
25—35	3	16	10	..	..	29
35—45	..	10	15	7	..	32
45—55	..	..	7	10	4	21
55—65	..	..	..	4	5	9
Total ..	9	29	32	21	9	100

(M. Com., Alld., 1944).

(27) (*a*) Discuss fully what is meant by the coefficient of correlation and how it is measured and interpreted.

(*b*) Calculate the co-efficient of correlation from the following:

Subject (Age of husband)	17	18	19	19	20	20	21	21	22	23
Relative (Age of wife)	12	16	14	15	15	19	22	16	15	20

(B. Com., Alld., 1942).

(28) What do you understand by co-efficient of concurrent deviations?

Calculate the coefficient of concurrent deviations from the following data :—

n or number of pairs of observations=47.
c or number of pairs of concurrent deviations=16.

(29) Calculate the co-efficient of concurrent deviations from the data given in exercise 22, page 437.

(30) Calculate the co-efficient of correlation between production of wheat and its price upon the basis of figures given in exercise 33, Chapter XI, page 222.

(M.A., Patna, 1941).

(31) Calculate the co efficient of correlation between infant mortality and overcrowding from the following data —

Infant mortality	109	122	96	142	151	121	125	102	109	156	122
Percentage of population over crowded	14.9	6.3	5.8	12	33.2	13.3	14.6	8.8	4.9	39.3	6.3

(B Com Hons Andhra 1943)

(32) From the following table calculate the co-efficient of correlation between the ages of husbands and wives and estimate the probable error of the result obtained

Age of husbands	Age of wives								Total
Years	15	20	25	30	35	40	45	50	
55—60						6	5	3	14
50—55					8	10	6	2	26
45—50				2	13	8	4		2
40—45				14	18	3			35
35—40			15	20	8				43
30—35	6	12	25	16					59
25—30	15	26	20	1					62
20—25	22	10	2						34
Total	43	48	2	53	47	27	15	5	300

(B Com Nagpur 1943)

CHAPTER XVIII

ASSOCIATION OF ATTRIBUTES

Statistics of Attributes.

In Chapter VIII[1] reference was made to descriptive and numerical characteristics of statistical data, and characteristics like sex, civil condition, caste, religion and infirmity were mentioned as descriptive. It will be well to note here again that quantitative character of statistical data arises in two ways. First, the investigator may note only the *presence* or *absence* of some attribute in a series of objects or individuals, and *count* the number of those who possess it and of those who do not. For instance, in a given population the number of the deaf and not-deaf, or of the sane and insane may be counted. In such cases, the quantitative character arises solely in the process of counting. Second, the investigator may note or measure the actual *magnitude* of some variable character for every one of the objects or individuals observed. For instance, height of students in a class, length of leaves of a certain tree, or prices of certain commodities may be *recorded*. Such records are quantitative in character. In these cases therefore the observations themselves are quantitative in character.

As pointed out in Chapter VIII, the first kind of observations are based on descriptive characteristics and are termed **Statistics of Attributes,** while the second are based on numerical characteristics and are called **Statistics of Variables.** So far, in all the chapters, we have been concerned with statistics of variables. We have studied how variables are analyzed, compared, and correlated with one another. It is now proposed to study how re-

---

[1] *Cf* pages 95-96.

lationship can be established between two attributes and how that relationship can be discovered by the method of Association

Notation and Terminology

While discussing classification of data according to attributes in Chapter VIII it was pointed out that when one attribute is noticed two distinct classes are formed. These two classes however arbitrary their boundary are exclusive of each other Such classification was in that Chapter, denominated as simple classification or classification by dichotomy

To discuss the theory of association and its application in practice it is necessary to have some simple notation for the classes formed and for the measurements assigned to each of them Accordingly, we shall use the capital letters A B C to denote the several **attributes.** An object or individual *possessing* the attribute A will be termed simply A that possessing B B The class whose members *possess* the attribute A will be termed the **class A.** Similarly we shall use the small letters *a b c* (generally, the Greek letters α β, γ are used) to denote the *absence* of the attributes A B C Thus if A represents the attribute blindness *a* represents sight i e non blindness if B stands for insanity, *b* stands for sanity **Combinations** of attributes will be represented by grouping together the letters that indicate the attributes concerned Thus if A represents blindness and B insanity AB represents blindness and insanity If the presence and absence of these attributes are noticed then

Combination	AB stands for	blindness and insanity
	A*b*	blindness and sanity
	*a*B	sight and insanity
	ab	sight and sanity

and similarly the class AB includes all those who are blind and insane the class A*b* all those who are blind and sane and so on. If a third attribute be noted for example deafness and

denoted by C, the class ABC includes those who are at once blind, insane and deaf, and the class AB*c* those who are blind and insane but not deaf.

The number of observations assigned to any class will be termed the frequency of the class, or briefly **class-frequency.** Class-frequencies will be denoted by placing the corresponding class-symbols in brackets. Thus,

(A) denotes number of A's, *i.e.*, objects possessing attribute A.

(A*b*) denotes number of A*b*'s, *i.e.*, objects possessing attribute A but not B, and so on.

The attributes denoted by capitals A B C ...... may be termed **positive attributes,** and their contraries denoted by small letters **negative attributes.** Thus the classes A, AB, ABC are positive classes; the classes *a, ab, abc,* negative classes. AB and *ab*, A*b* and *a*B, A*b* C and *a*B*c*, are pairs of **contrary classes.** A class specifying one attribute is known as the class of first order; while that specifying two, that of the second order. Thus, A is a class of the **first order,** AB or BC that of the **second order.** Similarly, (A), (A*b*), (*a*BC) are class-frequencies of the first, second and third orders respectively. The series of classes given by any one positive class and the classes whose symbols are derived therefrom by substituting small letters for one or more of the capital letters in all possible ways will be termed as **aggregate.** Thus (AB), (A*b*), (*a*B), (*ab*) form an aggregate of frequencies of the second order. When no attributes are specified, the total number of observations constitutes the **Universe** with its limits specified, and will be denoted by the letter N.

It should now be clear that the Universe must be equal to the number of A's plus the number of *a*'s Similarly the

number of A's should equal the number of A's that are B plus the number of A's that are not B, and so on It means that any class frequency can be analyzed into higher class frequencies Thus,

$$N=(A)+(a),$$
$$N=(B)+(b),$$
$$N=(AB)+(Ab)+(aB)+(ab)$$
$$(A)=(AB)+(Ab)$$
$$(B)=(AB)+(aB)$$
$$(a)=(ab)+(aB)=N-(A)$$

The classes specified by attributes of the highest order are termed the **ultimate classes** and their frequencies the **ultimate class-frequencies** If we know (AB) and (Ab) we can find (A) and if in addition we know (ab) and (aB) we can not find only find (a) but also N This is due to the fact noted above, that every class frequency can be expressed as the sum of certain of the ultimate class frequencies Therefore to specify the data completely it is only necessary to know the ultimate class frequencies An example will further clear the point

Example 1 Given the following ultimate frequencies find the frequencies of the positive and negative classes and the whole number of observations N —

$$(AB)=100 \quad (Ab)=50$$
$$(aB)=80 \quad (ab)\ 40$$

To find N we must remember that it will be equal to the aggregate of the frequencies of the second order Thus

$$N=100+50+80+40$$
$$=270$$

The frequency of any first-order class e g (A) is given by the total of the two second-order frequencies the class symbols for which contain the same letter Thus

$$(A)=(AB)+(Ab)=100+50=150$$
$$(B)=(AB)+(aB)=100+80=180$$
$$(a)=(aB)+(ab)=80+40=120$$
or $(a)=N-(A)=270-150=120$
$$(b)=(Ab)+(ab)=50+40=90$$
or $(b)=N-(B)=270-180=90$

	A	a	N
B	(AB) 100	(aB) 80	(B) 180
b	(Ab) 50	(ab) 40	(b) 90
N	(A) 150	(a) 120	N 270

The nine-squares table given above affords an easy and quick manner of getting the required class-frequencies. If given values are filled in it, the required ones may be computed from it, as shown in the table.

Similarly, if eight ultimate frequencies of the third order are given, or sixteen ultimate frequencies of the fourth order are given, all the positive and negative class-frequencies and N can be obtained from them by mere addition.

If the values of any two classes in each of the equations used is the solution of the above example are known, the value of the third can be easily found. For example,

if, $(a)=(aB)+(ab)$
then $(ab)=(a)-(aB)$.

And, the expression of any class-frequency in terms of the positive frequencies is most easily obtained by a process of step-by-step substitution; thus

$$\begin{aligned}(ab)&=(a)-(aB)\\&=[N-(A)]-[(B)-(AB)]\\&=N-(A)-(B)+(AB).\end{aligned}$$

The expression of other class-frequencies in terms of positive frequencies can be made by a similar process of substitution.

Probability and Expectation.

When a coin is tossed once, it must fall heads or tails. The probability (pure chance) that it would fall head is is $\frac{1}{2}$. When a coin is tossed 50 times, the expectation of a head coming up is $\frac{1}{2}\times 50=25$. Therefore **expectation** is equal to the product of probability and the number of observations. If two coins are tossed the chance of two heads or two tails coming up is reduced to $\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$.

If two attributes A and B are studied in a Universe N and the class frequencies of the attributes are (A) and (B).

$$\text{Probability of (A)} = \frac{(A)}{N}$$

$$\text{Probability of (B)} \quad \frac{(B)}{N}$$

$$\text{Probability of (A) and (B) Combined } \frac{(A)}{N}\times\frac{(B)}{N}$$

$$\text{And, Expectation of (A) and (B) combined} = \frac{(A)}{N}\times\frac{(B)}{N}\times N$$

$$= \frac{(A)\times(B)}{N}$$

Criterion of Independence.

When actual observation is equal to expectation attributes are **independent** and there is no association between them. In such a case we expect to find the same proportion of A's amongst the B s as amongst the non B s.

Let us take an example

Example 2. If (A) = people vaccinated — 50

(B) = people not attacked by small pox = 60,

(AB) = people vaccinated but not attacked — 20

N — total number of people = 150

it is required to find whether the attributes A (vaccination) and B (freedom from attack) are independent.

In this case, expectation of $(AB) = \frac{(A)\times(B)}{N}$

$$= \frac{50\times 60}{150} = 20.$$

The actual observation (people vaccinated but not attacked) is thus equal to the expecation. Therefore, A and B are independent. We conclude that vaccination and freedom from attack are not associated with each other in the given case.

Let us take another example:

Example 3. If in the above example (AB) are not given, but instead, we know the number of people who were not vaccinated and were attacked by small-pox, *i.e.*, (ab), to be equal to 60, we proceed to find whether the attributes a and b are independent.

Expectation of $(ab) = \frac{(a)\times(b)}{N}$

Now $(a) = N - (A) = 150 - 50 = 100$

and $(b) = N - (B) = 150 - 60 = 90$

Therefore, $\frac{(a)\times(b)}{N} = \frac{100\times 90}{150} = 60$

Again, the actual observation is equal to the expectation. In this case also the two attributes, a and b, are independent.

We can now put down the criterion of independence in more convenient form when actual class-frequencies of the second order only are given.

Attributes A and B are independent if (AB), actual observation,

$$= \frac{(A)\times(B)}{N} \text{ (expectation).}$$

Attributes a and b are independent if (ab), actual observation,

$$= \frac{(a)\times(b)}{N} \text{ (expectation).}$$

Attribute A and b are independent if (Ab) actual observation

$$= \frac{(A)\times(b)}{N} \text{ (expectation)}$$

Attributes a and B are independent if (aB) actual observation,

$$= \frac{(a)\times(B)}{N} \text{ (expectation)}$$

$$\text{Hence } (AB)\times(ab) = \frac{(A)\times(B)}{N} \times \frac{(a)\times(b)}{N}$$

$$= \frac{(A)\times(b)}{N} \times \frac{(a)\times(B)}{N}$$

$$\text{But } \frac{(A)\times(b)}{N} \times \frac{(a)\times(B)}{N} = (Ab)\times(aB)$$

Therefore $(AB)\times(ab)=(Ab)\times(aB)$

This last equation gives the required criterion of independence for the attributes A and B in the case of actual ultimate frequencies of the second order

We take one more example

Example 4 Let the actual observations be as follows —
People vaccinated but not attacked by small pox $=60=(AB)$
People not vaccinated and attacked by small pox $=272=(ab)$
People vaccinated but attacked by small pox $=80=(Ab)$
People not vaccinated and not attacked by small pox $=204=(aB)$

It is required to find whether the attributes A (vaccination) and B (freedom from attack by small pox) are independent.

According to the criterion just indicated A and B are independent only when

$$(AB)\times(ab)=(Ab)\times(aB)$$

Now in the given case

$$(AB)\times(ab)=60\times272=16320$$
$$(Ab)\times(aB)=80\times204=16320$$

The criterion is satisfied and therefore, the attributes A and B are independent that is not associated with each other

Association and Disassociation.

In statistics the word association has a technical meaning, distinct from the one current in ordinary speech. Ordinarily one speaks of A and B as being associated if they appear together in a number of cases. It is not so in statistics, where A and B will be said to be associated only if they appear together in a larger number of cases than is to be expected if they are independent. The mere fact that *some* A's are B's however great the proportion, is not enough to show that A and B are associated. This is a fundamental principle.

Association may be positive or negative. There is a simple way of knowing it. If two attributes, A and B, are not independent, but related to each other, then if

$$(AB) > \frac{(A) \times (B)}{N}$$

A and B are said to be **positively associated.** If, on the contrary,

$$(AB) < \frac{(A) \times (B)}{N}$$

A and B are said to be negatively associated, or, briefly, **disassociated.** It should be carefully noted that disassociation does not mean the same thing as independence.

In example 4, with the data as given, A and B are independent; that is, they are not related. If the actual observation of people who were vaccinated but not attacked by small-pox, that is, if class-frequency (AB) were more than 60, and the expectation were 60, the attributes would have been positively associated. But if the actual cases of (AB) were less than 60, while expectation remained 60, A and B would have been disassociated.

Upon the above principle we take an example.

Example 5. Let $(A)=40$; $(B)=35$; $(a)=10$; $(b)=15$; $(AB)=30$; $(Ab)=10$; $(ab)=5$; $(aB)=5$; and, $N=50$.

From the given observed frequencies we can construct a table like the following

Table X (Observation)

	A	α	N
B	30	5	35
b	10	5	15
N	40	10	50

We now find the expectations

Expectation of $(AB) = \frac{(A)\times(B)}{N} = \frac{40\times35}{50} = 28$

Expectation of $(\alpha b) = \frac{(\alpha)\times(b)}{N} = \frac{10\times15}{50} = 3$

Expectation of $(B) = \frac{35}{50}\times50 = 35$

Expectation of $(A) = \frac{40}{50}\times50 = 40$

And so on

We may now construct a table of expectations as below

Table Y (Expectation)

	A	α	N
B	28	7	35
b	12	3	15
N	40	10	50

If we compare table X with table Y we shall be able to study the fact of positive and negative associations In table Y, (AB)

is 28, while in table X it is 30. This implies that actual observation of (AB) is greater than its expectation, or in other words,

$$(AB) > \frac{(A) \times (B)}{N}$$

Therefore, vaccination and freedom from attack, to make A and B stand for what they did so far in our examples, are positively associated.

On the other hand, (Ab) in table Y is 12 and in table X it is 10. That is, actual observation of (Ab) is less than its expectation. Therefore, A and b are negatively associated. Or, in table Y aB is 7, while it is only 5 in table X. This implies that

$$(aB) < \frac{(a) \times (B)}{N}$$

Therefore, a and B are disassociated or negatively associated.

Co-efficient of Association.

So far we have ascertained the fact of association by comparing the observed class-frequencies with the expectations. We have not measured the degree of association. Several co-efficients have been devised for judging the intensity of association. Of these, the following co-efficient due to Yule is the simplest:

$$Q = \frac{(AB)\ (ab) - (Ab)\ (aB)}{(AB)\ (ab) + (Ab)\ (aB)}$$

where, Q stands for the Co-efficient of Association. This Co-efficient is zero when the attributes are independent, +1 if they are completely associated and —1 if they are completely disassociated.

Let us take an example.

Example 6 Required to compute the co-efficient of association from the data given in table X of example 5 above

$$Q = \frac{(30\times5)-(10\times5)}{(30\times5)+(10\times5)}$$
$$= \frac{100}{200}$$
$$= 5$$

Hence the intensity of association between the attributes A and B is 5 and the association is positive.

Let us take another example

Example 7 Required to compute the co efficient of association from the data given in example 1 above

$$Q = \frac{(100\times40)-(50\times80)}{(100\times40)+(50\times80)}$$
$$= 0$$

Hence the attributes A and B are independent

Yule s co-efficient of association is quite easy to compute and is a convenient measure of association since it not only exhibits the intensity of positive and negative associations but also shows the independent character of the attributes

Partial Association.

If in a given case it is found that

$$(AB) > \text{or} < \frac{(A)\times(B)}{N}$$

all that this information leads us to is that A and B are related with each other in some way We cannot say whether the relationship is direct or of any other kind. It is possible that association between A and B may not be direct, but due to the association of A with C and of B with C. An example will make the point clear

An association is observed between 'vaccination' and exemption from attack by 'small-pox', that is, more of the vaccinated people are exempt from attack than the unvaccinated ones. It may be argued that this does *not* imply that vaccination protects the people from attack, but that most of the unvaccinated people are drawn from those living in insanitary and filthy conditions. Thus A (vaccination) and B (exemption from attack are associated due to the association of both with C (hygienic conditions).

The ambiguity in the above case arises from the fact that the universe contains not only the objects possessing the third attribute alone, or objects not possessing it, but both. In our example, both hygienic and non-hygienic conditions may be prevailing in the locality where observations have been made. If, however, the universe of observation were confined to either class alone, for instance, the observations relating to vaccination and attack were made from a narrow section of the population living under approximately identical hygienic conditions, and still A and B were found to be associated, above ambiguity would not arise.

The associations found between the attributes A and B in the universe of C's and the universe of *c*'s are termed as **partial** associations, to distinguish them from total associations found between A and B in the universe at large.

Partial association may prove to be entirely misleading, for what is true of the whole is not true of each of the parts. To take our example again, observations regarding vaccination and attack may be drawn from people living under same hygienic conditions, *yet* some of the people may be rich and others poor. There may be a positive association between vaccination and exemption from attack among the rich but not among the poor. The disparity between these results may be explained by the simple fact that the poor are more open

to attack than the rich so that attack is not independent of poverty Or it may be that only the rich get themselves vaccinated and vaccination is in this case, not independent of poverty

Thus an illusory or misleading association may arise in a case where in the given universe there exists a third attribute C with which both A and B are associated positively or negatively If both the associations are of the same sign the resulting illusory association between A and B will be positive if of opposite signs the illusory association will be negative For example if the associations between A and C and between B and C are positive they would give rise to an illusory positive association between A and B

Illusory association may also arise in a different manner, that is through the personality of the observer If the attention of the observer fluctuates it is likely that he may observe the presence of A when he observes the presence of B and *vice versa*. In such a case A and B will both be associated with the observer's attention C and an illusory association will result.

EXERCISES

(1) How would you distinguish between 'association' and 'correlation' as the terms are used in statistics?

(M A Agra 1944)

(2) Given the following ultimate class frequencies find the frequencies of the positive and negative classes and the whole number of observations N —

$(AB)=200 \quad (Ab)=100$

$(aB)=160 \quad (ab)=80$

(3) From the following data find whether the attributes A and B are independent

$(A)=100 \quad (B)=120 \quad (AB)=40 \quad N=300.$

(4) $(AB)=120$, $(ab)=544$, $(Ab)=160$ and $(aB)=408$, find whether the attributes A and B are independent.

(5) Given the following ultimate class-frequencies, find the frequencies of the positive classes.

$(ABC)=298$ $(AbC)=450$ $(aBC)=408$ $(abC=342$
$(ABc)=1476$ $(Abc)=2292$ $(aBc)=3524$ $(abc)=43684$

(6) Given the following frequencies of the positive classes, find the frequencies of the ultimate classes:

$(N)=47{,}426$	$(ABC)=312$
$(A)=3{,}236$	$(AB)=856$
$(B)=4{,}030$	$(AC)=670$
$(C)=1{,}540$	$(BC)=312$

(7) Show whether A and B are independent, positively associated or negatively associated in the following cases:—

(1) $N=1000$ $(A)=470$ $(B)=620$ $(AB)=320$,
(2) $(AB)=512$ $(aB)=1536$ $(Ab)=96$ $(ab)=288$
(3) $(A)=245$ $(AB)=147$ $(a)=285$ $(aB)=190$

(8) Investigate the association between darkness of eye-colour in father and son from the following data.

Fathers with dark eyes and sons with not dark eyes	=237
Fathers with dark eyes and sons with dark eyes	=150
Fathers with not dark eyes and sons with dark eyes	=267
Father with not dark eyes and sons with not dark eyes	=2346

(9) Given the following data find whether deaf mutism and baldness are associated:—

Total population	..	16,264,000
Number of the bald-headed	..	24,441
Number of the deaf-mutes	..	7,623
Number of the bald-headed deaf-mutes	..	225

(10) Find the association between eye-colour of husband and eye-colour of wife from the following data:—

Husbands with light eyes and wives with light eyes	..	1236
Husbands with light eyes and wives with not light eyes	..	856
Husbands with not light eyes and wives with light eyes	..	528
Husbands with not light eyes and wives with not light eyes		476

(11) Can vaccination be regarded as a preventive measure for small pox from the data given below

Of 1,482 persons in a locality exposed to small pox, 368 in all were attacked.

Of 1,482 persons 343 had been vaccinated and of these only 35 were attacked

(M Com, All 1944)

(12) The following table gives the number of persons suffering from certain infirmities in Bengal in 1931 —

Sex	Total Number	Insane	Deaf mutes	Deaf mutes and Insane
Males	260 lakhs	12,6[illegible]0	21,301	545
Females	241	9,055	14,136	317

Trace the association between insanity and deaf muteness for males and females of Bengal separately

(M.A., Alld, 1938)

(13) (a) Write a short note on the use of Co-efficient of Association in analyzing economic statistics

(b) From the figures given in the following table compare the association between literacy and unemployment in rural and urban areas and give reasons for the difference, if any —

	Urban	Rural
Total Adult Males	25 Lakhs	200 Lakhs
Literate Males	10 Lakhs	40 Lakhs
Unemployed Males	5 Lakhs	12 Lakhs
Literate and Unemployed Males	3 Lakhs	4 Lakhs

(M A, Ald, 1937)
(M.A. Patna 1943)

(14) Find the association between inoculation against cholera and exemption from attack from the following data:

	Not attacked	Attacked	Total
Inoculated	276	3	279
Not inoculated	473	66	539
Total	749	69	818

(15 In the course of anti-malarial work quinine was administered to 606 adults out of a total population of 3,540. The incidence of malarial fever is shown below. Discuss the preventive value of quinine.

	Fever	No-fever	Total
Quinine	19	587	606
No Quinine	193	2,741	2,934
Total	212	3,328	3,540

(M.A., Cal., 1935).

(16) Criticize the following arguments:—

(1) 99 per cent. of the people who take alcohol die before they reach the age of 80 years. Therefore, taking alcohol is bad for longevity.

(2) 99 per cent. of the members who voted for the tenancy bill were cultivators. Therefore it was unfair to suppose that the voting was unbiassed,

(17) The male population of the U P is 250 lakhs. The number of literate males is 20 lakhs and total number of male criminals is 26 thousand. The number of literate male criminals is [illegible] thousand Do you find any association between literacy and criminality?

(M A Agra 1943)

(18) Investigate whether there is any association between extravagance in father and son from the following data —

Extravagant sons with extravagant fathers	496
Miser sons with extravagant fathers	162
Extravagant sons with miser fathers	184
Miser sons with miser fathers	1158

(18) Write explanatory notes on the following

Association Criterion of Independence, Partial Association, Dissassociation

CHAPTER XIX

INTERPOLATION AND FORECASTING

Meaning of interpolation.

Interpolation stands for the **insertion of the most likely estimate under certain assumptions.** In Chapter X, the mode and the median were interpolated in the modal and the median classes respectively; but, this was done only by starting with certain assumptions in both the cases. In locating the mode in a continuous frequency distribution it is assumed, as it was done in Chapter X, that mode is influenced by the class-intervals adjacent to the modal class[1]; while, in locating the median in a similar series it is assumed that the magnitude of the median class is uniformly distributed over its frequencies[2]. Location of the mode and the median in a grouped distribution suggests examples of interpolation as also the usefulness of this device for estimating some missing figure in a series.

Necessity of Interpolation.

In the absence of complete data at our disposal there would be no way out, except that of resorting to interpolation, to find the mode and the median. Hence the necessity of this method in such cases. But there are cases other than these where gaps may have to be filled in. Such gaps may be due to the fact that no record has been made, or its details are insufficient, or it has been lost or destroyed. Cases in point arise in connection with returns like those of the census which are, and can be, taken only once in a few years, so that

[1] *Cf* pages 122 and 125.
[2] *Cf* page 131.

if population figures are wanted for any intervening year, as they are in several instances an estimate has to be made of the most likely figures from the results already recorded. For example it may be necessary for purposes of administration or the like for a local or central government to be able to know with a reasonable degree of accuracy the population of an urban or rural area or a province at any given time, or to know the area under particular crops or the area under irrigation Similarly a sociologist an economist or a businessman may be interested in knowing the likely estimate of a certain phenomenon he is concerned with. A sociologist may like to know the number of people in different age-groups during the intercensal period an economist may desire to have a knowledge of the total tax revenue raised in a certain year, while a businessman may like to fill up the gaps in his yearly sales register In all these cases it cannot be supposed without any valid reason that the figures relating to a past year would apply to the year whose figures are required to be estimated Nor can mere imaginary figures be relied upon The most likely estimate has to be made.

Such an estimate may relate to some past date or to future one The technique of estimating a past figure is termed as **Interpolation** while that of estimating a probable figure for the future is called **Extrapolation.** To make an estimate certain assumptions are necessary

Assumptions

The first assumption that is made in interpolation or extrapolation is that there are no sudden jumps from one period to another If population figures for India for 1911, 1921 and 1941 are given and an estimate has to be made for the figure for 1931 this would be done only when it is assumed that there was no violent disturbance in the intermediate

dates, nor was the year 1931 an exceptional year such as that affected by epidemics, war or other calamity.

The second assumption is that in the absence of evidence to the contrary the rise or fall has been uniform. That is, in our example, the population growth has to be assumed to be uniform between 1921 and 1941, if the year 1911 or some other information has not to modify this assumption.

Accuracy of Interpolation.

Upon the above assumption figures may be interpolated, but the question that arises is: 'What is the certainty that the interpolated figures, which by hypothesis are unknown, are in reality the most probable figures?' In the words of Dr. Bowley, the accuracy of interpolation depends "(1) on knowledge of the possible fluctuations of the figures, to be obtained by a general inspection of the fluctuations at dates for which they are given; (2) on knowledge of the course of the events with which the figures are connected."[3]

It follows, therefore, that in basing arguments upon such figures the fact that they are interpolated ones should not be lost sight of. Interpolated figures are based on quite a different class of evidence from those figures which result from direct evidence. In some instances interpolations may represent figures which do not exist and which are used only for convenience of calculation. For instance, in allotting monthly marks to a student who was absent from a few seminars, attention may be paid to the student's general place in the class and to the average marks got by the students present in those particular seminars. Marks thus allotted have no existence. In other cases, interpolated figures may be, in the absence of the knowledge of full facts, the most probable estimates of figures that really exist. Therefore, all such estimates must be indicated as interpolations; it is always better to point the

[3] Bowley, A. L., *Element of Statistics*, 1920 ed., p. 217.

method by which they are obtained If any subsidiary information which may be regarded as a direct evidence of the accuracy of interpolated figures is available it is well to state it also. Further if practicable interpolated figures should be stated not as exact ones but as lying in a range within which their accuracy may not be questioned

Methods of Interpolation

Figures can be interpolated by the graphic method or by algebraic treatment Graphic method is good to follow when the quantities show cyclical character We shall discuss below with suitable examples the graphic method, the method of fitting a parabolic curve the method of advancing differences (Newton s method) and the Lagrange s formula.

The Graphic Method

Graphic Method in a Continuous Time Series.—This method may be explained by an example Table 63 gives the population of the province of Bengal during the last seven censuses Column 1 of the table shows the independent variable (x) which advances by an equal increament of 10 years Column 2 shows the corresponding values of the dependent variable (y)

Table 63 *Population of Bengal*
1881—1941

Year	Population in lakhs
x	y
1881	363
1891	391
1901	421
1911	455
1921	467
1931	501
1941	603

Suppose the figure 421 for the year 1901 is not given, and we are required to interpolate it. The task of interpolation would depend upon the evidence available for the purpose. If, for example, we know only the figures 391 and 455 relating to the years 1891 and 1911 respectively, we may plot them on a graph paper with years on the base line and population on the vertical scale. Since only a straight line would result from joining the points, we have no alternative but to assume that the population between 1891 and 1911 rises at a uniform rate. In the absence of any information to the contrary this is the most correct assumption possible. The height of the ordinate drawn from 1901 to intersect the straight line would give us an estimate of population of Bengal in 1901. This would be equal to $\frac{455+391}{2}$ or 423 lakhs, which exceeds the actual figure (421 lakhs) only by 2 lakhs. The difference is not very great, the error being of a little less than .5 per cent. If an estimate of population for any other year during the intercensal period is required, an ordinate from the particular year may be raised to intersect the straight line. The height of the ordinate on y would give the figure for the particular year.

If, on the other hand, figures are available for all the years excepting 1901, the various y's shall be plotted against their respective x's on a graph. If the resulting points are joined as straight lines with a ruler, we will have to assume sudden jumps in the growth of population, at least at the figures for the years 1911, 1921 and 1931, *i.e.*, at points C, D, and E in Figure 31. We have read that rather than making such an assumption, we should assume that in the absence of evidence to the contrary sudden changes in the quantities from one period to another do not occur. Therefore, instead of joining the points by straight lines we should draw through them all a line whose curvature is as smooth as possible. Such

a curve may be constructed on mathematical principles or drawn freehand. In Figure 31 it has been drawn freehand connecting each (x y) point. To find the y proper to 1901, we have drawn an ordinate through 1901 intersecting the curve at P. The height of this ordinate which is 421 gives the figure in lakhs for the population of Bengal in 1901. The remarkable closeness with which the interpolated figure agrees with the actual one given in table 63 is more or less accidental.

Interpolation of Population of Bengal by Graphic Method

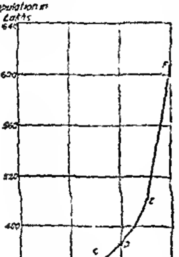


Fig 31

The principle followed in Figure 31 is not an unreasonable one to adopt for, in effect it gives due weight to each of the observations (y s) actually recorded and it assumes an even course from each year to the next—a quite justifiable assumption in the absence of any evidence that some sudden discontinuity or break has taken place in the y s.

Graphic Method and Periodic Figures.—If we have a series of monthly averages of figures relating to a certain phenomenon say sales of silver or price of wheat in India, and the

averages show periodic fluctuations, which we can study by the method discussed while dealing with Analysis of Time Series in Chapter XVI, we can interpolate figures for any month for which the records are incomplete. This can be done with a fair degree of accuracy. In general, the amount of sales of silver in India would show a rise during summer every year for that is the marriage season in the country, and the price of wheat would show a fall in April-May when new crop appears on the market. The curves drawn for such phenomena would exhibit a kind of periodicity, *i.e.*, they would regularly rise and fall. This would enable the fillng in of unknown figures in a manner which would not be unsatisfactory, For, if we know that a curve ought to rise or fall to a certain limit to be in conformity with its periodicity, we get a reliable clue to the position of the missing figures. Further, even those figures which lie at the two ends of our series of averages, and which probably cannot be found by any other method, can be traced out by the graphic method, when once the cyclic character of the curve has been known. We shall see later the usefulness of such curves for purposes of forecasting.

Graphic Method and Correlation Curves.—If, by the use of correlation graphs, discussed in Chapter XVII, we are able to find a close connection between two series, we can use one of them, whichever is more complete, to help in interpolating a missing figure in the other. First, we should carefully study the closeness between them at the dates for which we have complete figures in both series, and then draw a figure similar to Figure 30, one of the lines being, of course, incomplete. Thereafter, we may complete the incomplete line in such a manner that the complete line would be in close resemblance with the line that is already complete. Thus we shall obtain the most probable values for the figures which are missing in the incomplete series.

This method can be very usefully employed in interpolating figures for the values of exports from those of imports for the amount of money in circulation from figures of prices for the production of sugar in India from statistics of imports of sugar for changes in parts of the population from changes in the whole and for many other series only if we know that correlation of a high degree exists between them

Algebraic Treatment

The problem of interpolation to which greatest attention has been paid is as follows

If one quantity is subject to continuous regular change and a second quantity changes in connection with it and if we know or can estimate directly only some discontinuous value of this second quantity then it is required to estimate the most probable value of the second quantity corresponding to given values of the first For instance given the annual premia payable on a life policy at ages 25 30 etc. years it is required to interpolate the premium for intermediate ages or given the population of India in 1881 1891 1901 and 1911 it is required to interpolate it for intermediate dates or extrapolate it for future dates

Two assumptions as already noted are made in such cases. Firstly it is assumed that the quantity (premium or population in the above examples) changes continuously that is, without any sudden break at any figure And secondly it is assumed that the rate of change of the quantity is likewise continuous so that the curve representing it is smooth and not angular

The problem stated above can be tackled systematically by algebraic treatment. We take up below three of the several methods available for interpolation

First Method—Fitting with a parabolic curve.—To make the argument as general as possible we shall speak of x and y as variables and assume the value of y as depending on that of x in such a way that when x is given, y is known of can be estimated.

Suppose

$$y=a+bx+cx^2+ \ldots\ldots\ldots\ldots\ldots$$

where, a, b, c. . . . are constants to be determined, and their number can be made to depend upon the number of known values of y. The equation

$$y=a+bx+cx^2+ \ldots\ldots\ldots\ldots nx^n$$

represents a curve called a parabola of the n^{th} order.

Let us illustrate the method by fitting a parabolic curve to the following figures giving the population of Allahabad at decennial censuses:—

Year	1901	1911	1921	1931
Population in thousands	172	171.7	157.2	183.9

It is required to interpolate the population for 1916. Assuming that no abnormal conditions prevailed in 1916 to cause a sudden change in the population of Allahabad, let us proceed to estimate the population for that year with the help of the given data. Since the known points are four in this particular case, we would take as the curve through them a parabola of the 3rd order, *viz*,

$$y=a+bx+cx^2+dx^3 \ldots\ldots\ldots\ldots (1)$$

Then the four known points would be just sufficient to determine the four constants a b c d Now the x class intervals are equal being of 10 years each we measure from 1916 as origin and get

$x=$	-15	-5	0	$+5$	$+15$
$y=$	172	171 7	y_0	157 2	183 9

where y_0 is the number to be estimated

To further simplify the algebra we may take 5 years as unit for x so that

$x=$	-3	-1	0	$+1,$	$+3$
$y=$	172	171 7,	y_0	157 2	183 9

Since all five points are to lie on the curve with equation as in (1), we substitute the above values of x and in the equation and get

$$172=a-3b+9c-27d \qquad \text{(i)}$$
$$171\,7=a-b+c-d \qquad \text{(ii)}$$
$$y_0=a \qquad \text{(iii)}$$
$$157.2=a+b+c+d \qquad \text{(iv)}$$
$$183\,9=a+3b+9c+27d \qquad \text{(v)}$$

Equation (iii) tells us that y_0 the value that we have to interpolate is equal to the constant a Therefore from equations (i) (ii) (iv) and (v) we proceed to deduce the value of the constant a

Adding (i) and (v)

$$2a+18c=172+183\,9 \qquad \text{(vi)}$$

Adding (ii) and (iv),

$$2a+2c=171\,7+157.2$$

or $18a+18c=9\ (171\,7+157\,2$ (vii)

Subtracting (vi) from (vii),

$$16a = 9\ (171.7+157.2)-(172+173.9)$$
$$= 2960.1 - 355.9$$
$$= 2604.2$$

Therefore, $a = 162.7625$

or, $y_0 = 162.763$

The population of Allahabad, as interpolated, is 162.763 for 1916.

In a like manner interpolation of population for any other intercensal year or extrapolation for any year after 1931 can be made. It should be remembered that if data were given for three years, equation giving the parabola of the 2nd order, *viz*, $y=a+bx=cx^2$, would be used. Similarly, if data were given for five periods, the equation used would be that of the 4th order, and so on.

The above method results in simultaneous equations, like the equations, (i) (ii), (iv) and (v), and is also sometimes called the method of simultaneous equations.

Second Method—By means of advancing differences.— This method is also known as **Newton's method.** The *following figures, table 64, show the amount of annual premia* required by an insurance company to secure Rs. 1,000 without profits. It is required to calculate the amount of premium payable at the age of 22 next birthday.

Table 64 *Annual Premia on a life policy of Rs 1,000, and their differences*

Age next birth-day in years x	Annual premium in Rs y	Differences First Δ^1	Second Δ^2	Third Δ^3	Fourth Δ^4	Fifth Δ^5
20 x_0	25 y_0					
		3 Δ_0				
25 x_1	28 y_1		1 Δ^2_0			
		4 Δ_1		0 Δ^3_0		
30 x_2	32 y_2		1 Δ^2_1		.5 Δ^4_0	
		5 Δ_2		.5 Δ^3_1		-.25 Δ^5_0
35 x_3	37 y_3		1.5 Δ^2_2		.25 Δ^4_1	
		6.5 Δ_3		.75 Δ^3_2		
40 x_4	43.5 y_4		2.25 Δ^2_3			
		8.75 Δ_4				
45 x_5	52.25 y_5					

Each entry in the difference columns is formed by taking the *algebraic* difference of the entries on the left Thus,

$\Delta_0 = y_1 - y_0 = 28 - 25 = 3$; $\Delta_1 = y_2 - y_1 = 32 - 28 = 4$;

$\Delta^2_0 = \Delta_1 - \Delta_0 = 4 - 3 = 1$, $\Delta^2_1 = \Delta_2 - \Delta_1 = 5 - 4 = 1$,

$\Delta^3_0 = \Delta^2_1 - \Delta^2_0 = 1 - 1 = 0$; $\Delta^5_0 = \Delta^4_1 - \Delta^4_0 = .25 - .5 = -.25$

In this manner, differences have been calculated in main columns 3, 4, 5, 6 and 7 of the table.

The formula to be used for interpolating the value of y for a given x due to Newton is,

$$y_x = y_0 + x\Delta_0 + \frac{x(x-1)}{1\times 2}\Delta^2_0 + \frac{x(x-1)(x-2)}{1\times 2\times 3}\Delta^3_0$$
$$+\frac{x(x-1)(x-2)(x-3)}{1\times 2\times 3\times 4}\Delta^4_0\ldots\ldots$$

where, y is the figure to be interpolated, Δ's are the differences, and x is calculated as follows:

$$\frac{\text{Year of interpolation}-\text{Year of origin}}{\text{Time distance between adjoining years}},$$

The above equation uses differences upto the 4th order, and can be extended to the nth order.

It is required to know the amount of premium payable at age 22 years next birthday.

Now, from the table we find.

$$y_0=25;\ x=\frac{22-20}{25-20}=\frac{2}{5}=.4;$$

$\Delta_0=3$; $\Delta^2_0=1$; $\Delta^3_0=0$; $\Delta^4_0=.5$; $\Delta^5_0=-.25$.

By substituting the above values upto this—the fifth—order of differences,

$$y_x=25+(.4\times 3)+\frac{.4(.4-1)}{1\times 2}\times 1+\frac{.4(.4-1)(.4-2)}{1\times 2\times 3}\times 0$$
$$+\frac{.4(.4-1)(.4-2)(.4-3)}{1\times 2\times 3\times 4}\times .5$$
$$+\frac{.4(.4-1)(.4-2)(.4-3)(.4-4)}{1\times 2\times 3\times 4\times 5}\times -.25.$$

$$=25+1.2+\frac{-.24}{2}+0+\frac{-.9984}{24}\times .5+\frac{3.6}{120}\times -.25.$$

$$=25+1.2-.12+0-.0208-.0075$$

$$=26.05.$$

The required annual premium payable at age 22 years is Rs 26.05

Interpolation for the value of y corresponding to any other x can also be made in a like manner

Newton's formula uses differences running in a diagonal direction and is suited for interpolation near the beginning of the table It should be noted that it is used in a case in which the independent variable x advances by **equal increments**. In table 64 x advances by 5 years When it is required to interpolate in a frequency distribution it is better to work with the cumulative numbers For example if from the frequency distribution of marks given in table 22 page 195 it is desired to know the total number of candidates who obtained marks not exceeding 15 the table, for purposes of calculating the differences and *interpolating the number of students* would be written as follows (table 65), and interpolation carried out as above

Table 65 *Cumulative Frequency of Marks in Economics*

Number of marks x	No of candidates y
Not more than 10	4
" " " 20	12
" " " 30	23
" " " 40	38
" " " 50	49
" " " 60	56
" " " 70	60

Third Method—Lagrange's formula.—When the recorded y's correspond to x's, and the x's advance by **unequal intervals**, the most convenient formula to use is that due to the famous

French mathematician, Lagrange, known after his name as Lagrange's formula.

Representing the quantities as before by

$(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots\ldots\ldots\ldots\ldots\ldots (x_n, y_n)$,

Lagrange's formula runs thus:

$$
\begin{aligned}
y_x = {} & y_0 \frac{(x-x_1)(x-x_2)\ldots\ldots\ldots\ldots\ldots(x-x_n)}{(x_0-x_1)(x_0-x_2)\ldots\ldots\ldots\ldots\ldots(x_0-x_n)} \\
& + y_1 \frac{(x-x_0)(x-x_2)\ldots\ldots\ldots\ldots\ldots(x-x_n)}{(x_1-x_0)(x_1-x_2)\ldots\ldots\ldots\ldots\ldots(x_1-x_n)} \\
& + \ldots\ldots\ldots\ldots\ldots \\
& + y_n \frac{(x-x_0)(x-x_1)\ldots\ldots\ldots\ldots\ldots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\ldots\ldots\ldots\ldots\ldots(x_n-x_{n-1})}
\end{aligned}
$$

where, y_x is the quantity to be interpolated, x is the given quantity in the x variable corresponding to which y_x is to be interpolated, $x_0, x_1, x_2, \ldots\ldots\ldots x_n$ are the given values of the variable x, and $y_0, y_1, \ldots\ldots\ldots y_n$ are the corresponding given values of the variable y.

The following table relates to income earned per month by a certain number of workers in a big manufacturing concern.

Table 66. *Monthly Income of Workers.*

Income not exceeding Rs.	x	Number of peons	y
15	x_0	36	y_0
25	x_1	40	y_1
30	x_2	45	y_2
35	x_3	48	y_3

It is required to estimate the number of workers getting not exceeding Rs. 26 per month.

Making use of the given data and taking $x=26$, we have,

$$y_x = 36\frac{(26-25)\ (26-30)\ (26-35)}{(15-25)\ (15-30)\ (15-35)} + 40\frac{(26-15)\ (26-30)\ (26-35)}{(25-15)\ (25-30)\ (25-35)}$$

$$+45\frac{(26-15)\ (26-25)\ (26-35)}{(30-15)\ (30-25)\ (30-35)} + 48\frac{(26-15)\ (26-25)\ (26-30)}{(35-15)\ (35-25)\ (35-30)}$$

$$=36\times\frac{1\times-4\times-9}{-10\times-15\times-20} + 40\times\frac{11\times-4\times-9}{10\times-5\times-10}$$

$$+45\times\frac{11\times1\times-9}{15\times5\times5} - 48\times\frac{11\times1\times-4}{20\times10\times5}$$

$$= -\frac{54}{125} + \frac{792}{25} + \frac{297}{25} - \frac{264}{125}$$

$$= \frac{5127}{125}$$

$$=41$$

Therefore, income not exceeding Rs. 26 per month is being earned by 41 workers, as interpolated.

The above example relates to a case of frequency distribution where the magnitude of different class intervals is not equal. If, instead of frequency distribution, individual items were given as, for instance, population of India during certain years, the years advancing not necessarily by equal intervals, the method of attacking the problem would remain the same as that in the above example. In a like manner extrapolation can be made.

Forecasting

While discussing graphic method of interpolation used in connection with periodic figures it was pointed out on page 465 that when the cyclical character of a curve had been ascertained it was easy to locate a missing figure, and it was also hinted

there that such curves could prove useful for forecasting. Indeed they can, for it has been found that economic events move in a cycle. Periods of industrial boom or of agricultural depressions have been found to repeat themselves at an interval of 7—10 years. Therefore, when once it has been found, as a result of the study of sufficient and reliable data, that a certain phenomenon is characterised by cyclical tendency, its future course can be fairly accurately predicted upon the strength of past knowledge. This prediction is nothing but forecasting. Many businessmen can make forecasts about the future of their business without actually drawing a curve or even without knowing the name of periodicity. For instance, a chemist knows it well that malaria season is the one in which sales of quinine would be the largest during a year, and a bullion merchant always expects a rise in the price of silver during the months from March to June with the approach of marriage season in India. Thus, every business has its season. The practical businessman knows the facts of ups and downs in business from his experience; to him a periodicity curve or forecast based on it is of little interest. But to a statisician or to an economist the knowledge of periodicity is of great assistance in predicting the course of many economic events and taking advantage of them.

But even businessmen in the western countries are making use of what are called Economic Barometers. These barometers are special compilations made for the purpose of indicating tendencies of economic events. The construction of Indices of Business Conditions has already been explained in Chapter XII[4]. It has been pointed out there that business in general passes through well-defined minor and major changes, which fact makes business forecasting possible. How this

[4] *Cf*, page 257.

forecasting is actually done has been discussed in Chapter XIII[1] while dealing with the indices of business conditions prepared in England and the U S A The Harvard Committee of Economic Research publishes these indices in the form of charts These charts help business forecasting Similarly, Forecasting Composite Line published by the Brookmire Economic Service helps in forecasting stock and commodity prices in the U S A

Again in Chapter XV[2] while dealing with logarithmic or ratio scale charts it was pointed out that such charts, once their fluctuating character had been studied could be extended to predict a future figure relating to the phenomenon they illustrated This is again done with a knowledge of the trend of the curve. In Figure 21 page 354 the dotted line shows the manner of extending the curve beyond the last date upto which evidence is available This is a method of extrapolation which leads to forecasting From the extended portions of the curves in Figure 21 it can be predicted that the amount in 1943 would rise to Rs 5064 and Rs 506.4 respectively in the two cases

Conclusion.

After a discussion of the various methods of interpolation and extrapolation the practical utility of these methods must now be clear In administration as well as in business maintenance of minute records of each item from date to date is a matter of time labour and money It is impossible to take yearly census of population for instance But population figures are year after year necessary for estimating the income from existing taxes for exploring the possibilities of new sources and estimating the additional income How use-

---

[1] Cf pages 285 289
[2] Cf page 306

ful are the methods, discussed above, for estimating the population year after year can be very easily seen. Similarly, a businessman's sales records may be incomplete or he may like to estimate the probable demand for his wares in the coming year. Methods such as those discussed above would provide him with the most likely estimate based on past experience. So, these methods are of immense practical use.

EXERCISES

(1) What is interpolation? Explain its necessity by taking a few examples.

(2) What are the assumptions that are made in interpolating missing figures in a series? How far are interpolated or extrapolated figures to be relied upon?

(3) What is extrapolation? Show by taking a few examples how extrapolation leads to forecasting.

(4) What are Economic Barometers? How far do they help in forecasting economic events?

(5) Give a few examples of the use of Interpolation in Business Statistics.

(M. Com., Luck., 1942)

(6) Explain fully the process of interpolation by graphic method.

(7) Estimate the expectation of life at the age of 16 years, using the following data:

Age in Years	Expectation of life
10	35.4 yrs.
15	32.3 yrs.
20	29.2 yrs.
25	26.0 yrs.
30	23 2 yrs.
35	20.4 yrs.

(M.A., Agra, 1943).

(8) What is interpolation? Under what assumptions are the positions of Mode and Median interpolated in a frequency distribution?

Estimate the annual sales of cloth for 1935 from the following records of a cloth dealer —

Year	Sales of cloth in lakhs of yards
1920	250
1925	285
1930 1935	328
1940	444

(M Com, Alld, 1944)

(9) Interpolate the population of India for 1901 by graphic method using the following data —

Year	Population in millions
1881	253
1891	287
1901	?
1911	315
1921	319

(10) The following are the annual premiums required by the Bharat Insurance Co Ltd Lahore to secure Rs 1 000 with profits by making 20 payments in all What would be the premium payable at age 26 next birthday?

Age next birthday	20 payments	
	Rs	As
20	36	1
25	39	2
30	42	13
35	47	6

(B, Com, Agra, 1932)

(11) The following table gives the number of income tax assessees in the U. P.:—

Incomes not exceeding	Number of assessees
Rs. 2,500	7,166
Rs. 3,000	10,576
Rs. 5,000	17,200
Rs. 7,500	20,505
Rs. 10,000	21,975

Estimate the number of assessees with incomes not exceeding Rs. 4,000.

(M.A., Alld., 1944).

(12) The following are the annual premiums in a certain Life Insurance Co., for a policy of Rs. 500 payable at the death with an agreed bonus:—

Age next birthday	25	30	35	40	45
Annual Premium	24/10	27/11	31/9	36/6	42/5

Calculate the premium at age 36.

(M. Com., Luck., 1942).

(13) The following table gives the annual premia required to secure at death Rs. 1000 under whole life policies of a certain Assurance Co.

Age next birthday (in years)	Annual premium (in Rupees)
25	26
30	28
35	30
40	35
45	42
50	51

Compute the premium required at the age 36 next birthday.

(B. Com., Luck., 1942).

(14) The following table gives the quantities of a certain brand of tea demanded at prices noted against each Estimate the probable demand when the price is Rs 1 14 a lb

Price of Tea per lb. (in Rs)	1/4	1/8	1/12	2/-	2/4
Quantity demanded (in thousand lbs)	82.5	70.8	63.1	55.0	48.9

(M.A. Alld. 1942)

(15) The following table shows the annual rounded off value of production in a factory for the period 1925—1935 Estimate the missing value for 1930

1925	5 005	1931	5.347
1926	5 012	1932	5 516
1927	5 031	1933	5 733
1928	5 068	1934	6 004
1929	5 129	1935	6,335

(M.A., Cal, 1937)

(16) The following table gives the population of Lucknow at the time of last six censuses —

1881	2 53 729
1891	2 64 953
1901	2,56,239
1911	2 3[illegible] 332
1921	2,17,167
1931	2 51 097

Estimate the population of Lucknow in 1941 by graphic method only

(B Com. Luck. 1938)

(17) Discuss the utility of interpolation and extrapolation to a businessman What are the different methods known to you for interpolation?

Interpolate the figures for 1921 by the algebraic method of finite differences :—

Year			Population of India
1901	..	..	294,261,056
1911	..	..	315,156,396
1921	..	..	?
1931	..	..	351,523,045

Test the validity of your method if you know the actual census figures for 1921.

(M. Com., Alld., 1943).

(18) State Newton's formula for interpolation. Calculate the sale of silk in 1928 from the following da a.—

Year			Total sales (in Rs. 1,000)
1927	..	..	233
1929	..	..	391
1931	..	..	582
1933	..	..	799
1935	..	..	1035

(M.A., Cal., 1936).

(19) What do you understand by interpo'ation and extrapolation? What are their uses? The following table gives the normal weights of babies during the first twelve months of life:

Age in months	0	2	5	8	10	12
Weight in lbs.	$7\frac{1}{2}$	$10\frac{1}{4}$	15	16	18	21

Estimate the weight of the baby at the age of 7 months.

(M.A., Patna, 1940).

(20) Interpolate to find the missing number in the following table:

Year	Index No. of Exports
1919	72
1920	57
1921	?
1922	81
1923	103

State the assumptions underlying the formula employed.

(B. Com., Nagpur, 1944).

(21) Estimate the probable number of persons earning between Rs 40 and 50 from the following data —

Income in Rs	Number of persons
Below Rs 20	120
Rs 20—40	145
Rs 40—60	200
Rs 60—80	250
Rs 80—100	150

2/0-50

(22) Interpolate the missing figures in the following table of rice cultivation —

Year	Acres in millions
1911	76.6
1912	78 7
1913	?
1914	77 7
1915	78 7
1916	?
1917	80 6
1918	77 6
1919	78 7

(B Com Agra 1937)

(23) The following table gives the census of population of certain town in 1891 1901 1911 1921 and 1931 Estimate the population in 1925 making your method clear —

Years	Population
1891	38 724
1901	132 785
1911	168 076
1921	195 690
1931	246,050

(M A Cal 1937)

(24) The following are the marks obtained by 492 candidates in a certain examination :—

Not	more	than	40	marks	212	candidates
"	"	"	45	"	296	"
"	"	"	50	"	368	"
"	"	"	55	"	429	"
"	"	"	60	"	460	"
"	"	"	65	"	481	"
"	"	"	70	"	490	"
"	"	"	75	"	492	"

Find out the number of candidates who secured more than 42 but not more than 45 marks.

(M.A., Cal,, 1935).

(25) From the data given in exercise 32, Chapter XI, page 222, estimate the number of persons whose income is Rs. 19 but does not exceed Rs. 25.

(26) The annual sales of a concern are given below.—

Years	Sales of Cloth in lakhs of yards.
1915	125
1920	163
1925	204
1930	238
1935	282

Assuming the conditions of the market to be the same, estimate the sales for the year 1940.

(M.A., Patna, 1941).

(27) The following table gives the population of Index at the time of the last six census:

1881	75,401
1891	82,984
1901	86,686
1911	44,947
1921	93,091
1931	127,327

Estimate the population for 1941.

(B. Com., Agra, 1944).

CHAPTER XX

INTERPRETATION OF DATA

The Science of Statistics, as we have defined it in Chapter II[1], is concerned with the study of the methods of collection, analysis and interpretation of quantitative data So far, we have been dealing with the various statistical methods employed in the collection and analysis of numerical facts It is proposed to deal here with the interpretation of statistical data

Interpretation.

Interpretation stands for the technique of drawing out inferences from an analytical study of the collected figures. While discussing the methods of collection, presentation comparison, correlation, interpolation, etc., we have put down wherever necessary, in appropriate places the essential precautions which must be kept in view in handling statistical facts for analysis as well as for commenting on the results of analysis. A repetition of all those cautions here is obviously uncalled for But it must be said that commonsense is as much a chief requisite and experience as great a teacher in the delicate task of interpretation as they are in collection and analysis of quantitative data If this fundamental principle is ignored fallacious inferences would be drawn, which would recoil on the statistician and his science The statistician, to repeat what we have said in Chapter II[2], is not an alchemist expected to produce gold from any worthless material, he is rather like a chemist capable of assaying the value the material contains and of extracting nothing more than this value

---

[1,2] Cf page 17

Freedom from bias and prejudice on the part of the statistician is, no doubt, necessary in collection and analysis of data; it is all the more so in interpretation, because it is interpretation more than—or, rather than—collection and analysis of data with which the layman is concerned. The *power* which figures carry with them is such that the layman can be as easily *impressed* by them as he can be *deceived*. From the advertiser's gallery, from the electioneering platform, from the propagandist's forum, from the partisan press and from a hundred other sources the man in the street is bombarded with tendentious figures put forward to support some *ex parte* statement. Sometimes such figures are reasonably and justifiably used to form a basis for the arguments built upon them; more often, they give an exaggerated picture of the truth, which may be due to ignorance or inadvertence, but has also been found to be influenced by motive, by deliberate wish to mislead. The layman is not unaware of this fact. If he distrusts all arguments based on figures, his attitude is like that of a reasonable man, who has not the training for himself to separate the wheat from the chaff, and is, therefore, inclined to suspect everything. Statistical methods are most dangerous tools in the hands of the inexpert. A taxi driven by one who does not know the art of driving might fall in a ditch or collide against other vehicles resulting in serious injury to the passers-by, the taxi and the driver. Taxi must therefore be driven by one expert in driving. So must statistics be handled by one who is expert in the subject. Most often what happens is that attracted by the power of impression which statistics command so many men are led to use them without knowing their limitations, and feel jubilant if they win their point. At other times, it happens that when the data have been scientifically collected and analyzed they fall into the hands of such people for purposes of interpretation

as hardly know the subject. These people sometimes under the impress of their preconceived notions and at other times due to their habit of criticising every thing they come across, indulge in hair-splitting as if they alone know the art of interpreting statistical data. It is therefore, established that if the task of interpretation is to be scientifically performed, it must be done by a true statistician who is above all prejudice and has the daring to call a spade a spade.

Preliminaries to Interpretation.

Before starting on interpretation the statistician should examine whether

(1) the data are adequate to base his judgment upon;

(2) the data are homogeneous and comparable;

(3) the data are properly collected and are without biassed errors;

(4) the data have been scientifically analyzed and all disturbing factors considered.

After satisfying himself on these preliminary points he should draw out reasonable inferences. Most of the mistakes that are made in interpreting figurative data arise from false generalisation, a few examples of which may now be considered.

Mistakes due to False Generalisation.

Let us suppose that an argument runs as follows:

The prices of agricultural commodities in India in 1943 were five times the prices in 1931. India's prosperity in 1943 had consequently increased by leaps and bounds.

The argument as it stands looks sound. Let it be agreed that the ratio of prices between the two years is correctly stated. Now 1931 was a year of depression, when agricultural commodities were indeed selling at very low prices and 1943

a year of war-boom. Therefore, the two periods are different, and allowance must be made for this difference before right comparison is possible. Again, are the prices of agricultural commodities a measure of India's prospedity? Let it be supposed that they are a measure of the prosperity of agriculturists in India. But, is the whole of India agricultural, or only a part of it is so? Evidently, the whole of India is not. Then, is what is true of a part necessarily true of the whole? The answer is in the negative. Further, even supposing that the income of agriculturists in 1943 was much more than what it was in 1931, does that measure their prosperity? The agriculturists are to spend also, and if in spending they pay six times more on the same items in 1943 than what they did in 1931, do they retain money with them or lose? They do the latter. And, what is prosperity —the mere income, or the surplus income? Lastly, what is the significance of 'leaps and bounds'? 'Leaps and bounds' may be an impressive term, but is meaningless to the statistician unless he knows the bounds of 'leaps and bounds'. These querries will suffice to understand what false generalisation may lead to, and in what direction the mind of a statistician should work to arrive at correct inference.

Let us take another example. It may be argued that the production of foodstuffs in a country in a certain year was only 5 per cent. less than that in the previous year. There was, therefore, no *real* food shortage in the year under reference. Again, the argument creates an impression. But let us analyze it. Is the number of mouths to be fed in the country in the particular year the same as it was in the previous year, or has it increased? If it has increased, the demand for foodstuffs, other things being equal, is expected to increase. Again, were the foodstuffs exported from the country in the previous year? And, have they been exported in the year under consideration? If they were not exported in the pre-

vious year and have been exported in the particular year, or if the exports in the particular year exceed the exports in the previous year the shortage of 5 per cent would increase to a higher figure so far as actual consumption is concerned. Further, were foodstuffs imported into the country in the previous year? If yes have they been imported in the particular year under study? If not there would be shortage for purposes of consumption And if the imports in the year under consideration are much below those in the previous year shortage for consumption would result These points would serve to make it clear that the task of interpretation is not strewn with roses it needs an analytical approach

Another example may be found in the argument that since the quantity and value of goods imported into a certain country have been increasing the country is prospering Supposing the figures of imports are correct the question arises how much of the imported goods are being re-exported? Suppose all are being retained within the country Then, is the consumption of goods made in the country increasing, constant or decreasing? If it is increasing or constant the *per capita* consumption of foreign and native goods is increasing and if increased *per capita* consumption is a measure of prosperity, the country is prospering But *per capita* consumption would increase remain constant or decrease according to the change that takes place in the population of the country If the country is being colonized and the increase in the number of immigrants far outweighs the proportionate increase in the imports it is difficult to say that the prosperity of the country is increasing Again if the consumption of native goods population remaining the same decreases the imports may be just sufficient to recompense this decrease and therefore *per capita* consumption may not increase There would then be hardly any increase in prosperity But if the

imports are in excess of the decrease in the consumption of native goods, the chances are that *per capita* consumption, and with it the prosperity, would increase. Again, let us look at the problem from another angle. If, along with increasing imports, the consumption of luxury goods made in the country is also increasing, it is difficult to conclude that *per capita* consumption is increasing. Luxury goods are consumed by the classes and not by the masses, but the classes may form only a small portion of the country's population. What is true of them is not necessarily true of the *whole* country. If the increment in their prosperity is nullified by the loss in the prosperity of the other section, national prosperity would not increase. These lines of thought would show the amount of care and caution required in interpreting quantitative data. They also show that arguments, though seemingly correct. may be highly fallacious, if they are not properly sifted and analysed.

Wrong Interpretation of Index Numbers.

We have already dealt with the limitations of averages in Chapter X,[3] and indicated there the fallacious conclusions to which a careless interpretation of averages might lead. Also, with regard to the use of both the general price and the cost of living indices we have given adequate caution in Chapter XII.[4] We may here take an example of wrong interpretation of index numbers of prices. If a general index series shows a rising tendency in a country, it may be argued that since price level, as indicated by the index series, is rising, there is inflation in the country. The argument here runs from effect to cause. This manner of arguing things is rather questionable and unsound. An effect may be the

[3] *Cf.* page 168.
[4] *Cf.* page 258.

result of a multitude of causes To single out one cause from this multitude, without valid reasons and corroboratory evidence is not justifiable Moreover, index number reveals a change in the relative value of the standard of deferred payments and of goods in general—the two sides of the quantity theory equation It is not to be inferred that a change in the level of prices is *necessarily* due to causes directly touching the value of the standard rather than to causes touching the value of the goods, which are being compared with the standard The index number *merely* reveals the change in relationship the cause of the change is another question Therefore, it is not safe to say that if an index series shows a rising tendency the cause for rise in price level is the increase in quantity of money pushed into circulation. Index number simply shows the tendency

Wrong Interpretation of Coefficient of Correlation.

Co-efficient of correlation simply shows that two variables are related to each other The value obtained for the co efficient in a certain case should be interpreted with great caution Suppose a negative correlation is found to exist between area under jute and area under rice in Bengal The argument may run that the cultivation of jute in Bengal is increasing at the expense of rice This *might imply* that the people of Bengal want more jute for manufacturing jute cloth, gunny bags, sand bags etc. than rice for direct consumption This implication may prove *incorrect, if, on* further investigation, it is found that increase in area under jute is due to war emergency requiring jute manufacturers, or due to relatively higher prices of jute or due to such climate changes in the province as favour the growth of jute more than that of rice, and the cultivation of rice has gone down at the same time either because of rice plots having gone

under a crop other than jute, or because of the relative fall in price of rice or of rice-growers having joined the armed forces of the country. So, although the area under jute is increasing while that under rice is decreasing, it does not necessarily mean that jute is being grown at the expense of rice. From the above line of arguments it follows that the interpretation of the value of co-efficient of correlation needs not only caution, but also a thorough knowledge of the facts that constitute other hypothesis.

Wrong Interpretation of Co-efficient of Association.

While dealing with association of attributes in Chapter XVIII[5] we made reference to partial association. We pointed out the reasons to which illusory association may be due, and the fallacious conclusions to which such an association might lead. One more example may be taken.

It is observed, at a general election to the provincial legislatures, that a greater proportion of the Hindu Mahasabha candidates; who spent more money than their opponents, the Congress candidates, won the election than the Congress candidates who spent less. It is argued that the Hindu Mahasabha candidates won because of their having spent more than the Congress candidates. That is, there is a positive association between "spending more than the opponent" and "winning." The argument would be sound only if, on further investigation, it is found that these two attributes are not influenced by a third attribute. If a third attribute influences them, the co-efficient of association would work out to be a high figure even though "winning" and "spending more than the opponent" are not related. For instance,

---

[5] *Cf.* page 452.

the policy principles and programme of the Hindu Mahasabha may have generally carried the day, and 'spending more' had nothing to do with its success

General directions for Interpretation*.

In all the above examples we have not doubted the accuracy of the data. We have rather supposed that the data are correct, properly collected and presented But, even with correct data we have seen how wrong and fallacious conclusions might be drawn. It will be seen that in all these cases what appeared to be correct at first sight was not *necessarily* so when further investigations were made Therefore, a statistical conclusion must be based on all possible investigations In other words statistical results should not be considered as the sole determinants of the value of given data. Statistical treatment affords only one method of judging a phenomenon It is not the only method available. Therefore, conclusions based on statistical analysis mean nothing more than what figures imply A statistician cannot be dogmatic about his conclusions. He cannot, and should not assert that his figures tell that a certain result *must* be such and such. It may be, it may not be It will be only when it is confirmed by other methods of studying the same phenomenon. This great limitation on the interpretation of statistics should therefore be always kept in view Again statistical laws are true only on an average or in the long run. *They* study the norm, and not the abnormality Statistics deals with the group and not the individual. These facts must not be ignored while interpreting and using statistical data.

* Read in this connection "Limitations and Distrust if Statistics" in Chapter III pages 27-32.

EXERCISES

(1) What kind of mistakes are generally made in interpreting statistical data? Give examples.

(B. Com., Alld., 1936).

(2) What conclusions would you draw regarding the economic activities of the people living in the U. S. S. R. (Russia) from the study of figures given in the following table:—

(1928=100)

	1929	1930	1931	1932	1933	1934	1935
Industrial Production ..	126	164	203	231	250	300	369
Out of Investment goods	131	185	240	279	307	382	481
Out of Consumers' goods ..	122	147	172	190	200	230	274
Net Imports ..	92	111	116	74	37	24	25
Net Exports ..	114	128	100	71	61	52	45

(B. Com., Alld., 1939).

(3) What do you understand by interpretation of Statistics? How is it done?

(4) How would you interpret the following table giving age-distribution per thousand of the population in 1911?

Age	Germany	France	England	U.S.A.	Japan	India
Under 10	234	171	209	222	244	276
10—20	203	166	190	198	198	192
20—30	164	158	173	187	154	178
30—40	139	148	152	146	138	142
40—50	105	127	115	106	101	99
50—60	76	104	80	72	77	61
60—70	51	77	51	43	57	36
70 & over	23	49	30	26	31	16

(5) Interpret the facts contained in the following table:—

	Population in 1931 in millions	Mean density per sq. mile	No. of females per 1000 males
India	352.8	195	940
Bengal	50.1	646	924
Madras	46.7	328	1025
U. P.	48.4	456	90[illegible]
Bihar & Orissa	37.6	454	1005
Bombay	21.9	177	901
Ajmer	0.5	207	892
Assam	8.6	157	900

(6) The following is an extract from an office draft of an Annual Report of a large Public Library. Recapitulate the essential statistical information in the form of a tabular statement and bring out impressively the comparison attempted in the Report.

Lending habits among borrowers vary from year to year. Topical events leave their impress on the number of borrowers and more particularly on the damage inflicted on books borrowed.

Whilst in 1938 only 15,000 books were lent out, the stress of events in 1939 attracted no fewer than 3,80,000 borrowers. These latter indented as many as 25,500 and either lost or damaged 800 books. In 1938 there were only 1,20,000 borrowers, 48,000 borrowing 4,000 books dealing with Section F (Illustrated News).

In Section I (Travel) 1,000 were lent out in 1938 but the number increased to 7,000 in 1939 while the number of borrowers increased from 2,000 to 10,000 and the losses diminished from 5 to 2 or by 60 per cent.

Section D is made up of Pamphlets. In 1938 issues amounted to 2,000 books and borrowers 25,000. The statistics only one year later were 3,000 books and 60,000 borrowers with 100 losses. In Section D 40 books were lost in 1938 while in Section F, 680 losses among 2,80,000 borrowers were recorded in 1939. It may be noted that in this section the issues in 1939 were 14,000.

Section C (Biography) was in 1938 responsible for 1,000 books, 2,000 borrowers and 2 losses, but in 1939 although the books had increased in number by exactly one half, the number of borrowers remained exactly the same as before, and so also the number of losses.

t

In Section B (Science) there was no increase over the 3,000 in 1938 but 15,000 borrowers in 1938 increased by 3,000 in 1939 The losses were exactly double, 3 in 1938 and 6 in 1939.

Curiously enough in Section A (Fiction) the number of books and books lost which were respectively 4,000 and 20 in 1938, were reduced in 1939 to exactly one-half of these numbers. The number of borrowers also diminished from 28,000 to 10,000 in 1939. The total number of books in 1938 was 170 of which 100 were recorded in Section F.

(M. Com., Luck., 1942)

(7) What inference do you draw regarding the Indian Business Activity from the following indices of "Capital" Index of Indian Industrial Activity?

August	1939	110.1
September	1939	117.4
October	1939	113.0
November	1939	111.9
December	1939	121.3
January	1940	116.6
February	1940	116.9
March	1940	112.1
April	1940	117.2
May	1940	122.0
June	1940	115.8
July	1940	115.7

(8) What are statistical fallacies? Give examples mentioning the factors responsible for them.

(9) Comment on the following conclusions:—

(1) Population of Kanpur doubled during the decade 1931-41. Therefore, the birth-rate for the town has also doubled.

(2) The export of gold from India is increasing. The people of India are, therefore, getting poorer.

(3) The national income per head in India has now increased to Rs. 100 from Rs. 30 in 1900. Therefore, the people of India are now more happy.

(4) The income from stamp duties has been increasing in India Therefore, the number of suits filed in courts is increasing

(10) How would the present World War influence.

(a) Population Census of 1951 in India
(b) Marriage rate in India
(c) International Comparison of Statistics,
(d) Life Insurance

(11) It is observed that intelligent fathers have intelligent sons and intelligent grandfathers have intelligent grandsons Therefore intelligence is hereditary—Comment?

(12)

Months 1941 42	Notes in Circulation (Crores of Rs)	Bombay Wholesale Price Index No
April	249	122
May	255	123
June	260	127
July	257	140
August	258	144
September	266	145
October	274	152
November	284	162
December	304	180
January	328	184
February	349	194
March	357	197

Calculate the co-efficient of correlation from the above data and fully discuss whether the co-efficient indicates that the rise in wholesale prices at Bombay is due to inflation

(13) Plot the index numbers of commodity prices in both Bombay and Calcutta (given in exercise 1? Chapter XVII page 432), against time in graph paper Study the graph attentively, and write a short review of the course of commodity prices in the two cities during the period May 1942 to February 1943

(M.A Agra 1944)

APPENDIX I A.

SPECIMEN OF A BLANK-FORM.

The following blank-form was used by the Central Bureau of Economic Intelligence, United Provinces, in an enquiry into family budgets of mill workers in the United Provinces.

I. FOOD

Article	Quantity.	Cost.
	Mds.	Rs.
Wheat		
Wheat Flour		
Gram		
Gram Flour		
Birra (*Bejhar*)		
Rice		
Barley		
Maize		
Juar		
Bajra		
Dal urd		
Dal Arhar		
Dal Mung		
Dal (......)		
Ghee		
Oil (......)		
Milk		
Sugar		
Gur		
Meat		
Fish		
Eggs		
Potato		
Other Vegetables		
Salt		
Spices		
Sweetmeats		
Fruits		
Tea		
Total		

II. FUEL.

Article.	Quantity.	Cost.
	Mds.	Rs.
Firewood		
Coal		
Dung-cakes		
Total		

III. LIGHT

Article.	Quantity.	Cost.
Kerosene oil		
... . oil		
Matches		
Total		

IV. HOUSE RENT

Rent.	Repairs	Total

F 32

CLOTHING AND FOOTWEAR

Item	No	Cost	Life	Cost per month
		Rs. a. p.	Months	Rs. a. p.
(a) MEN				
1 [illegible]				
2 Pyjama				
3 Shirt				
4 Saluka				
5 Waistcoat				
6 [illegible]				
7 Underwear				
8 Khesa (Lah)				
9 Napkin				
10 Rumal				
11 Socks				
12 Shoes				
13 Chappals				
14 Safa				
15 [illegible]				
16				
Total M				
(b) WOMEN				
17 Sari				
18 Pyjama				
19 Lahanga				
20 Shirt				
21 Saluka				
22 Orhni				
23 Burka				
24 Chappal				
25 Stockings				
26				
Total W				
(c) CHILDREN				
27 Dhoti				
28 Sari				
29 Lahanga				
30 Pyjama				
31 Shirt				
32 Saluka				
33 Orhni				
34 Shoes				
35 Chappal				
36 Cap				
37				
Total C				

VI. HOUSEHOLD REQUISITES.

Including receipts from home.

Article.	No.	Cost	Life	Cost per month
		Rs. a. p.	Months	Rs. a. p.
1. *Charpai*				
2. Re-netting				
3. *Dari*				
4. *Kathri*				
5. *Razai*				
6. Sheets				
7. Blanket				
8. Utensils				
9. Tinning				
10. Umbrella				
11. Mattresses				
12. *Hugga*				
13.				
Total				

VII MISCELLANEOUS

Item	Rs. a. p.
1 Sweeper	
2 Barber	
3 *Dhobi*	
4 Soap	
5 Hair Oil	
6 Medicine	
7 Education	
8 Conveyance	
9 Travel	
10 Tobacco	
11 *Pan Supari*	
12 Intoxicants (	
13 Recreation	
14 [illegible]	
15 Remittances	
16 Postage	
17 Subscription	
18 Newspaper	
19 Litigation	
20 Interest	
21 Debt	
Total	

VIII SUMMARY

Food	Fuel	Light	Clothing	House rent	Misc.	Total Exp.	Total Income	Savings (+ or −)	D[illegible]

APPENDIX I B.

SPECIMEN OF QUESTIONNAIRE FOR THE CONSUMPTION BUDGET OF AN ARTISAN'S FAMILY

A. PRELIMINARY.

1. Names of the village, *pargana, tehsil* and district.
2. Nearest post office, police station and railway station.
3. Name of the head of the family.
4. Religion and caste.
5. Number of members in the family residing with the wage-earner.
   Adults: men and women.
   Children (under 12): boys and girls.
6. Number of dependents not living with the wage-earner.
   Adults: men and women.
   Children (under 12): boys and girls.
7. Number residing outside the village and contributing to family income.

B. MONTHLY FAMILY INCOME.

1. Normal monthly earnings of the head of the family.
2. Similar earnings of other members of the family in the village.
3. Contribution to family income by those residing abroad.
4. Other sources of income.
   Kind.
   Cash.

C EXPENDITURE ON FOOD (MONTHLY)

Quantity and Cost of

1 Rice
2 Wheat flour
3 Barley *Jowar* Gram etc to be specified
4 Pulses to be specified
5 Oil ( )
6 *Ghee*
7 Salt and spices
8 Vegetables
9 Fruits
10 Meat and Fish
11 Sugar and *gur*
12 Tea.
13 Others

D EXPENDITURE ON FUEL AND LIGHT. (MONTHLY)

Quantity and Cost of

1 Coal or cow dung
2 Charcoal
3 Firewood
4 Kerosene oil
5 Castor oil
6. Others

E. EXPENDITURE ON RENT. (MONTHLY)

1 Is the house own or taken on rent?
2 If taken on rent the amount paid as rent
3 If own,
(a) Cost of repairs paid to labourers and suppliers of materials
(b) Was family labour used? If yes to what extent?
(c) Cost of white washing
(d) When was the house constructed cost of construction, life of the house?

F. EXPENDITURE ON CLOTHING AND FOOTWEAR (MONTHLY)

For each article under this head answer

1. number of articles purchased,
2. the period they are estimated to last,
3. total cost incurred when purchased,
4. estimated cost per month.

For Men :—

1. *Dhoties.*
2. *Pyjamas.*
3. *Kurtas.*
4. Shirts.
5. *Pagri,* turbans or caps.
6. Coats and waistcoats.
7. *Sherwanis.*
8. *Mirzais or bandis.*
9. *Dhusa* or Lohi.
10. *Angocha* or handkerchief.
11. Socks.
12. Shoes.

For Women :—

1. *Lahanga* or *Sari.*
2. *Phariya, Urhni.*
3. *Kurti.*
4. Bodice.
5. Petti-coat.
6. *Chadar* or *Burqa.*

For Children :—

1. *Dhoties.*
2. *Saries.*
3. *Kurta.*

4 Cap
5 Shirt
6 Bodice
7 *Pyjama* or *Lahanga*
8 Shoes
9 *Ingo la*

G EXPENDITURE ON HOUSEHOLD REQUISITES (MONTHLY)

Under this head also answer for each article

1 the number purchased
2 the period they are estimated to last
3 the total cost incurred when purchased
4 estimated cost per month

For bedding purposes —

1 *Charpai*
2 Bedding proper
Dari chadar gadda tikaf pillows pillow-cases blankets *rajais*
3 Carpet for floor

For utensils —

1 *Thali Parat*
2 *Degchi Batua Patili*
3 *Karhai Tawa*
4 *Chamcha Chimta*
5 Other

H EXPENDITURE ON MISCELLANEOUS ITEMS. (MONTHLY)

1 Amount paid to barber
2 washerman
3 sweeper

4. Amount paid to village *purohit* or *mulla*.
5. „ „ „ for religious functions
6. „ „ „ medical fees and medicine
7. „ „ „ education
8. „ „ „ travelling by rail and road
9. „ „ „ conventional necessities *Pan Supari Bhang*, tobacco, etc.
10. Interest on debt.
11. Repayment of debt.
12. Payment to married daughter
13. Payment to dependents not living in the village
14. Expenditure on amusements
15. „ „ litigation.
16. Any other expenditure, *e g*, on jewellery or ornaments, sending letters, *Tika-bindi*, etc.

ABSTRACT OF EXPENDITURE.

Family Income.

Family Expenditure.—

Food.

Fuel and light

Rent.

Clothing and footwear.

Household requisites.

Miscellaneous.

Balance of Income over expenditure (+ or —)

APPENDIX II

LIST OF IMPORTANT STATISTICAL PUBLICATIONS

(A) Indian

I Publications of the Department of Commercial Intelligence and Statistics

1. Indian Trade Journal (Weekly)
2. Accounts relating to the Sea borne Trade and Navigation of India (Monthly)
3. Monthly Statistics of Cotton Spinning and Weaving in Indian Mills
4. Monthly Statistics of the Production of Certain Selected Industries of India
5. Accounts relating to the Inland (Rail and River borne) Trade of India (Monthly)
6. Monthly Statement of wholesale prices of certain selected articles at various centres in India
7. Accounts relating to the Sea borne Trade and Navigation of India (Annual)
8. Statistical Abstract for India (Annual)
9. Agricultural Statistics of India —
   Vol I—India (Annual)
   Vol II—Indian States (Annual)
10. Estimates of Area and Yield of Principal Crops in India (Annual)
11. Indian Tea Coal Rubber and Coffee Statistics (published separately) (Annual)
12. Joint Stock Companies in India and in some Indian States (Annual)

13. Statistical Tables relating to Banks in India (Annual).
14. Statistical Tables relating to the Co-operative Movement in India (Annual).
15. Review of the Trade of India (Annual).
16. Large Industrial Establishments in India (Biennial).
17. Live-stock Statistics, India (Quinquennial).
18. Quinquennial Report on the average yield per acre of principal crops in India.
19. Crop Forecasts of Rice, Wheat, Cotton, Linseed, Sesamum, Groundnut, Sugarcane, Castorseed (periodically), (Also published in the Indian Trade Journal).

II. Reports of Committees and Commissions.

1. Datta's Report on the Rise of Prices in India (1912).
2. Report of the Economic Enquiry Committee (1925).
3. Report of the Royal Commission on Indian Agriculture (1928).
4. Report of the Taxation Inquiry Committee.
5. Industrial Commission Report.
6. Report of the Royal Commission on Indian Labour.
7. Banking Inquiry Committees Reports (Central and Provincial).
8. Reports of the Committees and Commissions on Indian Currency and Exchange.
9. Industrial Surveys in various districts of U.P.
10. Labour, Unemployment and Textile Enquiry Committee Reports (Provincial).
11. Tariff Board Reports.
12. Report of Bowley-Robertson Committee.

III Other Government Publications

1. Gazette of India (Weekly)
2. Provincial Gazettes (Weekly)
3. Labour Gazette Bombay (Monthly)
4. Central and Provincial Governments Budgets (Annual)
5. Administration Reports of Provincial Governments (Annual)
6. Administration Report of Railways in India (Annual)
7. Report of the Controller of Currency (Annual)
8. Census Reports (for India Provinces and Native States) (Decennial)
9. Working Class Family Budgets
10. Monthly Survey of Business Conditions in India
11. Guide to Current Official Statistics
12. Indian Labour Gazette (Monthly)

IV Non official Publications.

1. Sankhya (Journal of the Indian Statistical Institute) (Calcutta)
2. Capital (Calcutta) (Weekly)
3. Indian Journal of Economics (Allahabad)
4. Commerce (Bombay) (Weekly)
5. Indian Year Book (Times of India Bombay) (Annual)
6. Wealth of India by Wadia and Joshi
7. Wealth and Taxable Capacity by Shah and Khambata
8. India's National Income by V. K. R. V. Rao
9. Eastern Economist (Weekly)

(B) Foreign

I Publications of Great Britain

1. Board of Trade Journal and Commercial Gazette (Weekly)
2. Ministry of Labour Gazette (Monthly)

3. Journal of the Royal Statistical Society, London (Quarterly).
4. Annual Statistics Abstract of the United Kingdom.
5. Guide to Current Official Statistics of the United Kingdom.
6. Census Reports.
7. Census of Production Reports.
8. Reports of Commission on National Debt and Taxation.
9. London and Cambridge Economic Survey.

II. Publications of the League of Nations Geneva.

1. International Statistical Year Book (Annual).
2. Memorandum on Currency and Central Banks.
3. Memorandum on Public Finance.
4. Methods of Compiling Cost of Living Index Numbers (1925).
5. Methods of Conducting Family Budget Enquiries 1928.
6. Year Book of Labour Statistics (Annual).
7. International Labour Review.

III. Some Standard Books on Statistics.

1. King, W. I.—Elements of Statistical Methods.
2. Boddington, A. L—Statistics and their Application to Commerce
3 Connor, L. R.—Statistics in Theory and Practice.
4. Jones, D, C.—A First Course in Statistics.
5. Secrist, H.—Introduction to Statistical Methods.
6. Bowley, A. L—Elementary Manual of Statistics.
7. Bowley, A L—Elements of Statistics.
8. Zizek—Statistical Averages.
9. Yule, G. U.—Theory of Statistics.

10 Yule G U and Kendall—An Introduction to the *Theory of Statistics*

11 Mills F C—Statistical Methods applied to Economics and Business

12 Elderton W P—Frequency Curves and Correlation.

13 Fisher R A—Statistical Methods for Research Workers

14 Fisher—Making of Index Numbers

15 Davis and Nelson—Elements of Statistics

16 Richardson—Business Forecasting

17 Wheldon H J—Business Statistics and Statistical Method

18 Tippet L. H C—The Methods of Statistics

19 Florence—Statistical Methods in Economics

20 Rhodes—Elementary Statistical Methods

21 Davies and Crowder—Methods of Statistical Analysis

22 Kelley—Statistical Methods

23 Davenport—Statistical Methods

24 Thurstone L L—The Fundamentals of Statistics.

25 Elderton and Elderton—Primer of Statistics

APPENDIX III.

MEASUREMENT OF THE NATIONAL INCOME OF INDIA

SUMMARY OF THE SCHEME RECOMMENDED BY THE BOWLEY-ROBERTSON COMMITTEE

Dr. A. L. Bowley and Mr. D H. Robertson were invited by the Government of India to consider, *inter alia,* the materials available for estimating the national income and wealth of India. They submitted their report entitled 'A Scheme for an Economic Census of India, in 1934, wherein, stating that these materials were very defective, they put forward certain practical proposals for estimating the total national income of India.

'The national income,' according to the committee, 'is the money measure of the aggregate of goods and services accruing to the inhabitants of a country during a year, including the net increments to, or excluding net decrements from, their individual or collective wealth.'

Two methods of calculation have been pointed out: the first comprizing an evaluation of the goods and services accruing, and the second a summation of individual incomes. The first is the census of products method and the second is the census of incomes method. The first method is unlikely to be ever applicable over even the whole of the industrial field in India. Special caution in combining the results of the two methods may be necessary.

The census of products method involves—

(i) evaluating the *net* output of agriculture, mining, industry, and other productive enterprises at the point of production. Precaution is necessary to avoid double counting (*e.g*, counting both the output of wheat and the labour of the cattle employed in raising it);

(ii) adding the value added by transporting and merchanting agencies in the country to home produce goods and to imports

(iii) adding exercises on home produced goods and customs duties on imports

(iv) adding the value of imports (c i f) including gold and silver

(v) deducting the value of exports (f o b), including gold and silver

(vi) deducting the value of goods home produced or imported which are used for maintaining fixed capital, or stocks of raw and finished goods intact

(vii) adding the value of personal services of all kinds,

(viii) adding the yearly rental value of houses whether rented or occupied by the owners

(ix) adding the increments in the holdings of balances and securities abroad by individuals or by government, or deducting the decrement in such holdings, similarly deducting the increment in such holdings in the country by residents abroad, or adding their decrement

The method described above is the more fundamental of the two methods of evaluating the National Income Certain precautions in following the census of incomes method must be observed in order that the results arrived through it may tally with those obtained by the first method

(i) All self consumed produce and receipts in kind must be included in the individual's income valued at their selling price at the place of production. So must be the yearly value of houses occupied by the owners

(ii) All interest payments must be deducted before writing down individual's income

(iii) The incomes of all individuals in the country should be entered gross, *i.e.* before payment of direct taxation. To this total should be added the undistributed profits of companies and the net profits of Government enterprises. From this total should be subtracted the interest on Government loans other than for productive enterprises and the pensions of all ex-Government servants.

(iv) To the total so far reached should be added receipts from customs and excise, stamp duty and local rates.

The suggestions which Dr. Bowley and Mr. Robertson make (stated below) relate to the estimate of the **broad sections of National Income:** the various adjustments indicated above would have their place in final calculations of the National Income.

The investigation they propose for estimating the national income is primarily on the basis of production but a minor part depends on individual incomes. The proportion to be thus estimated is probably greater in the towns in India. They say, "partly owing to difference in nature of the products and partly because different methods of investigation are necessary, rural income is distinguished from urban income."

For *rural* income they recommend an estimate of the quantity and value of all goods and services arising from the land or rendered in the village, by the method of **intensive surveys in selected villages.**

For *urban* income they suggest, in the first instance, **surveys of the larger towns** based on a sample enquiry of the personnel and occupations of families, an estimate of their incomes by personal statements and by investigation of wages and salaries prevailing in the town. For incomes over Rs. 1,000 or at least over Rs. 2,000, income-tax statistics can be of much help.

They have also recommended an intermediate Urban Population Census

These three enquiries would be supplemented by a *census of Production* applied to factories using power mines and some other industries

Rural Surveys

The method advocated for selecting the villages for the purpose of intensive survey is that of *random* sampling It consists of making a list of all the villages in a province in geographical order of districts, and after deciding the number will be investigated marking out starting from some random number the required number of villages all nearly equally spaced. Thus every unit in the aggregate will have an equal chance of being taken in When a village has been once selected it should on no account be substituted by another

The report gives the following table which shows the proposed minimum number of village to be investigated in each province

Province	Number of Villages in Province	Number in Sample
Bengal	86 000	250
Bihar and Orissa	83,000	300
Bombay	21 000	200
Central Provinces	40 000	200
Madras	51 000	200
Punjab	35,000	200
United Provinces	106 000	300

To arrive at the total for India some estimates must be made for Assam N W F Province tea plantations of Bengal areas of Bengal where coal mining is important and the areas affected by earthquake and not completely resettled by the time of the enquiry

The invesigator should be trained and live in each village for 12 months. In many cases the villages could be grouped in threes or fours, say 30 miles of each other. To each of such groups a superior investigator would be attacched, who would live in the largest village and do supervision work. Each province should be under the charge of a qualified statistician; and the entire survey should be controlled by the Director of Statistics, whose appointment the committee recommend as part of the Permanent Staff. The necessary schedules should be prepared by the Director in consultation with Provincial Statisticians. They should be adapted to local conditions, and local terms of weights, measures, etc. should be used. The main enquiry would, no doubt, be directed to income, production, consumption and allied topics, yet the investigators would have ample time to report on subjects like health, co-operation, debt, etc.

Urban Surveys.

Random sample of towns is not recommended. The problem is to be dealt with step by step, first by a synchronous survey of those cities in which Universities can organise satisfactory investigation, secondly by making similar, though not so intensive, surveys of other towns. After the Rural Surveys and the University City Surveys are completed, efficient investigators should be engaged to survey selected towns.

University City Surveys.

In the organisation of these surveys central control should be combined with local autonomy. A central committee should be set up to draw up an outline schedule of enquiry, to advise on any points referred to it and to present a report on the whole subject in the end.

If the surveys fall to Government Colleges, co-operation of Director of Public Instruction and the Education Depart-

ment would be necessary If they fall to self governing Univer sities arrangements would be made with the Economics Depart ment of the Universities concerned. City survey should be directly carried on by one of the Economics staff The detailed investigations should be carried out by graduates or post graduates reading Economics

There are two methods of approach 1 Occupational 2 By families

1 An occupational census is almost essential In each industry and important occupation in the town enquiries should be made regarding current rates of earnings and wages estimated over the year and allowing for seasonal variations It would include not only those employed in constructive industries but also clerks municipal and railway employees tonga drivers and all others working for salaries or wages or making small profits The method of payment (piece or time) may also be recorded

2 An accurate list of houses or tenements is necessary Big towns say of more than 150 000 persons may better be *divided into wards or groups of wards so that a unit may consist* of 30 000 houses About 1000 houses may be selected in each unit on random basis and visited by investigator House once selected should not be substituted by another

The visitor should establish friendly relations with the residents Thereby he would be able to obtain reliable information about numbers sex age and occupation of the family group. Repeated visits may be necessary Schedules should be filled in immediately after and not during the visit

The totals should in case of doubt be given as within a certain range All existing data relating to the subject of the survey emanating from Central and Local authorities trade organisations etc should be studied. Co-operation of these official and non official organisations should be sought

Census of Production.

This census would be imposed by a special Act of the Central Legislature, making the communication of facts demanded compulsory. It would be conducted by the Director of Statistics. Industries employing 20 or more persons and using mechanical power, some small workshops, certain industries where mechanical power is not used such as brick-making and carpet manufacture, railways and all establishments under the Mines Act would be covered.

Progress of factory industry is, to some extent, at the cost of cottage industry. Therefore, it would be good if the two could be brought in relation to each other. If some yearly data regarding them could be procured, an idea of their relative increase or decrease would be available. The necessary facts to be collected are the aggregate value of the sales and the aggregate cost of materials for each factory. The difference approximately shows the national income, accruing to the factory, and when all factories are considered the aggregate difference minus depreciation of plant and change in value of materials and finished goods would be a measure of the contribution to the national income of the industry.

The classification of products should be the same as that of exports and imports. The employees should be classified as salaried persons and wage-earners, young and adult with a statement of the age division between the two sexes. Besides, details can be obtained of the amounts and values of different commodities produced, and of materials bought, and of power used.

The opposition, objections and difficulties which the investigator will encounter will be great, but with the periodic repetition of the census they would automatically decrease.

APPENDIX IV.

LOGARITHM.

Logarithm of a given number to the base ten is the power to which the base ten should be raised to equal the given number

$10000 = 10^4$	Logarithm of	$10000 =$	4
$1000 = 10^3$	" ,	$1000 =$	3
$100 = 10^2$	" ,	$100 =$	2
$10 = 10^1$	" "	$10 =$	1
$1 = 10^0$	" "	$1 =$	0
$.1 = 10^{-1}$	" "	$.1 =$	-1
$.01 = 10^{-2}$	" "	$.01 =$	-2
$.001 = 10^{-3}$	" "	$.001 =$	-3
$.0001 = 10^{-4}$	" ,	$.0001 =$	-4

From the above it will be seen that the log of 1 is 0 and log of 10 is 1. Therefore log of any number between 1 and 10 would be greater than zero but less than 1. That is it would be equal to 0+ a fraction. Similarly, log of any number between 10 and 100 is 1+ a fraction. Thus a logarithm may consist of two parts the whole number, and the fraction. The whole number part *e.g.* 0 or 1 in the above instances, is termed **characteristic** and the fractional part is termed **mantissa.**

To determine the characteristic of any given number, the following two rules should be noted

(1) If the given number is greater than one, the characteristic is always *positive* and is obtained by the formula $(n-1)$, where n stands for the number of significant digits before the decimal point

(2) If the given number is less than one, the characteristic is always negative and is obtained by the formula $(N + 1)$,

where N stands for the number of zeroes after the decimal point but before any significant digit. The minus sign of the negative characteristic is written at the top of the characteristic and not prefixed to it. Thus, the characteristic of minus two is written as 2 and not as —2.

In accordance with the above two rules characteristics of a few numbers are given below:

Number	Value of n (rule 1)	Characteristic
4539	4	3
453·9	3	2
45·39	2	1
4·539	1	0

Number	Value of N (rule 2)	Characteristic
·4539	0	$\bar{1}$
.04539	1	$\bar{2}$
.004539	2	$\bar{3}$
.0004539	3	$\bar{4}$

Characteristic for any number can be similarly calculated.

For calculating mantissa for different numbers, Logarithmic Table has to be consulted. Logarithmic Table giving the mantissa for any number having less than four digits is given at the end of this Appendix. Mantissa of the required number should be read out from this table irrespective of the position of the decimal. If the given number is composed of four or more digits, it must first be approximated to 3 digits. Then its mantissa should be read out from the table. If we are to find the logarithm of 4539 we first approximate it to three digits. The approximation of 454 is .6571. Therefore, logarithm of 4539 is 3.6571.

Two facts about mantissa should be clearly noted. Firstly, it is always positive irrespective of the fact whether the characteristic is negatve or positive. Secondly, mantissa is not affected by the position of the decimal point in the

logarithm of each number is found out and added. The antilog of the sum is the required product.

$$\text{Log. } (a\times b)=\text{Log. } a+\text{Log. } b$$
$$\therefore a\times b=\text{Antilog. } [\text{Log. } a+\text{Log. } b]$$

The logarithm of any number a divided by b is the difference of the logs. of a and b Therefore, when one number is divided by another, the logs. of both the numbers are found out and the antilog of the difference between the two logs. gives the desired quotient.

$$\text{Log. } \frac{a}{b}=\text{Log. } a-\text{Log } b.$$

$$\therefore \frac{a}{b}=\text{Antilog. } [\text{Log } a-\text{Log.} b]$$

The logarithm of any number raised to n^{th} power is n times the log. of that number. Therefore, when any given number is raised to any power, the logarithm of the given number is found and multiplied by the power to which the number has been raised. The antilog. of the product gives the value of the given number raised to the desired power.

$$\text{Log. } a^n=n \text{ Log. } a$$
$$a^n=\text{Antilog. } [n \text{ Log. } a]$$

The log. of a given number to the n^{th} root is equal to the log. of that number divided by n Therefore, when the value of any given number to any given root is desired to be obtained, the log. of the given number is found out. This log. is divided by the given root to which the given number is to be reduced. The antilog. of the quotient gives the value of the given number reduced to the desired root.

$$\text{Log. } \sqrt[n]{a}=\frac{1}{n} \text{ Log. } a$$

$$\therefore \sqrt[n]{a}=\text{Antilog.} \left[\frac{\text{Log. } a}{n}\right]$$

MATHEMATICAL TABLES

INSTRUCTIONS.

Table of Logarithms—This table gives the *mantissa* of figures. To find the *mantissa* of any given number from the table, the number should be approximated to three digits. *Mantissa* of a number is the same regardless of the position of the decimal point in it.

Table of Antilogarithms—This table gives the antilogarithms of the *mantissa* portion of any given logarithm. The position of the decimal point in the required figure should be determined on the basis of the characteristic of the given logarithm.

Table of Squares—In this table upto 316 one zero, and from 317 onwards two zeroes, are omitted in each square. If in the given figure the decimal point moves by *one* digit to the left, then the decimal point moves by *two* digits to the left in the square.

Table of Square Roots—This table gives two square roots for each number. For *odd digits* in the given number, the upper figure, and for *even* digits the lower figure should be taken. If in the given number the decimal point moves by *two* digits to the left, then the decimal point moves by *one* digit to the left in the square root.

Table of Reciprocals—If in the given number the decimal point moves by *one* digit to the right, then the decimal point moves by *one* digit to the left in the reciprocal

	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
		6138	6149	6160	6170	6180	6191	6201	6212	6222

	0	1	2	3	4	5	6	7	8	9
42	·6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	·6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	·6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	·6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	·6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	·6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	·6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	·6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	·6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	·7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	·7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	·7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	·7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	·7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	·7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	·7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	·7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	·7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	·7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	·7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	·7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	·7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	·8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	·8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	·8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	·8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	·8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	·8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	·8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	·8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	·8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	·8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745

	0	1	2	3	4	5	6	7	8	9
75	·8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	·8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	·8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	·8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	·8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	·9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	·9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	·9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	·9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	·9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	·9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	·9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	·9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	·9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	·9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	·9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	·9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	·9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	·9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	·9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	·9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	·9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	·9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

	0	1	2	3	4	5	6	7	8	9
00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285
·11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377
·14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183

	0	1	2	3	4	5	6	7	8	9
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234
·35	2239	2244	2249	2254	2259	[illegible]	2270	2275	[illegible]	[illegible]
·36	2291	2296	2301	2307	2312	[illegible]	[illegible]	[illegible]	2333	2339
·37	2344	2350	2355	2360	2366	2371	[illegible]	2382	2388	2393
·38	2399	2404	2410	2415	2421	2427	[illegible]	2438	2443	2449
·39	2455	2460	2466	2472	2477	[illegible]	[illegible]	2495	2500	[illegible]
·40	2512	2518	2523	2529	2535	2541	[illegible]	2553	2559	2564
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685
·43	2692	2698	2704	2710	2716	2723	[illegible]	2735	2742	2748
·44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560
	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667

	0	1	2	3	4	5	6	7	8	9
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152
·79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528
·98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750
·99	9772	9795	9817	9840	9863	9886	9908	9931	[illegible]	[illegible]

SQUARES

	0	1	2	3	4	5	6	7	8	9
10	1000	1020	1040	1061	1082	1103	1124	1145	1166	1188
11	1210	1232	1254	1277	1300	1323	1346	186[illegible]	139[illegible]	1416
12	1440	1464	1489	1513	1538	1563	1588	1613	1638	1664
13	1690	1716	1742	1769	1796	1823	1850	1877	1904	1932
14	1960	1988	2016	2045	2074	2103	2137	2161	2190	[illegible]220
15	2250	2280	2310	2341	2372	2403	2434	2465	2496	[illegible]528
16	2560	2592	[illegible]674	2657	2690	2723	275[illegible]	[illegible]789	2822	[illegible]856
17	[illegible]890	[illegible]924	29[illegible]8	2993	3028	3062	3098	3133	3168	3[illegible]04
18	3 40	3 6	3312	3349	3386	34[illegible]3	3460	3497	3534	3572
19	3610	3648	[illegible] 6	3725	3764	3803	384[illegible]	3881	3920	3960
20	4000	4040	4080	4121	4162	4203	4244	4285	4326	4368
1	4410	445[illegible]	4494	4537	4580	4623	4666	4709	4752	4[illegible]96
[illegible]	4840	4654	4928	4973	5018	5063	5108	5153	5198	5244
[illegible]3	5[illegible]90	5336	5382	54[illegible]9	5476	5523	55 0	561	5664	5712
24	5760	5808	5856	5905	5954	6003	605[illegible]	6101	6150	6[illegible]00
[illegible]5	6[illegible]50	6300	6350	6401	6452	6503	6554	6605	6656	6[illegible]08
[illegible]6	[illegible]760	681[illegible]	6864	6917	6970	70[illegible]3	[illegible]0 6	7129	7182	[illegible]236
[illegible]7	7[illegible]90	7344	[illegible]398	7453	7508	7563	[illegible]618	7673	7728	7[illegible]84
[illegible]	7840	[illegible]896	952	8009	8066	8123	8180	823[illegible]	8294	8352
[illegible]	8410	8468	8526	8585	8644	8[illegible]03	876[illegible]	88[illegible]1	8880	8940
30	9000	9060	91[illegible]0	9181	9[illegible]4[illegible]	9303	9361	9425	9486	9548
31	9610	9672	9734	979	9860	99[illegible]3	9986		—	—
31		—						1005	1011	1018
3[illegible]	10[illegible]4	1030	1037	1043	1050	1056	106[illegible]	10[illegible]9	1076	1082
33	1089	1096	110[illegible]	1109	1116	1122	11[illegible]	1136	1142	114[illegible]
34	1156	1163	1170	1176	1183	1190	119[illegible]	1204	1211	1218
35	1225	1232	1239	1246	12[illegible]3	1 60	126	1274	12[illegible]	12[illegible]
36	1 96	1303	1310	1318	13 5	1332	1340	134[illegible]	1354	1362
3	1369	13 6	1384	1391	1399	1406	1414	1421	1429	143[illegible]
38	1444	1452	1459	1467	1475	148[illegible]	1490	1498	150[illegible]	1513
39	1521	1529	153	1544	1552	1560	1568	1576	1584	159[illegible]
	)	1608	1616	1624	1632	1640	1648	1656	1665	1673

	0	1	2	3	4	5	6	7	8	9
41	1681	1689	1697	1706	1714	1722	1731	1739	1747	1756
42	1764	1772	1781	1789	1798	1806	1815	1823	1832	1840
43	1849	1858	1866	1875	1884	1892	1901	1910	1918	1927
44	1936	1945	1954	1962	1971	1980	1989	1998	2007	2016
45	2025	2034	2043	2052	2061	2070	2079	2088	2098	2107
46	2116	2125	2134	2144	2153	2162	2172	2181	2190	2200
47	2209	2218	2228	2237	2247	2256	2266	2275	2285	2294
48	2304	2314	2323	2335	2343	2352	2362	2372	2381	2391
49	2401	2411	2421	2430	2440	2450	2460	2470	2480	2490
50	2500	2510	2520	2530	2540	2550	2560	2570	2581	2591
51	2601	2611	2621	2632	2642	2652	2663	2673	2683	2694
52	2704	2714	2725	2735	2746	2756	2767	2777	2788	2798
53	2809	2820	2830	2841	2852	2862	2873	2884	2894	2905
54	2916	2927	2938	2948	2959	2970	2981	2992	3003	3014
55	3925	3036	3047	3058	3069	3080	3091	3102	3114	3125
56	3136	3147	3158	3170	3181	3192	3204	3215	3226	3238
57	3249	3260	3272	3283	3295	3306	3318	3329	3341	3352
58	3364	3376	3387	3399	3411	3422	3434	3446	3457	3469
59	3481	3493	3505	3516	3528	3540	3552	3564	3576	3588
60	3600	3612	3624	3636	3648	3660	3672	3684	3697	3709
61	3721	3733	3745	3758	3770	3782	3795	3807	3819	3832
62	3844	3856	3869	3881	3894	3906	3919	3931	3944	3956
63	3969	3982	3994	4007	4020	4032	4045	4058	4070	4083
64	4096	4109	4122	4134	4147	4160	4173	4186	4199	4212
65	4225	4238	4251	4264	4277	4290	4303	4316	4330	4343
66	4356	4369	4382	4396	4409	4422	4436	4449	4462	4476
67	4489	4502	4516	4529	4543	4556	4570	4583	4597	4610
68	4624	4638	4651	4665	4679	4692	4706	4720	4733	4747
69	4761	4775	4789	4802	4816	4830	4844	4858	4872	4886
70	4900	4914	4928	4942	4956	4970	4984	4998	5013	5027
71	5041	5055	5069	5084	5098	5112	5127	5141	5155	5170
72	5184	5198	5213	5227	5242	5256	5271	5285	5300	5314
73	5329	5344	5358	5373	5388	5402	5417	5432	5446	5461

	0	1	2	3	4	5	6	7	8	9
74	5476	5491	5506	5520	5535	[illegible]	[illegible]	[illegible]	5595	5610
75	5625	5640	5655	5670	5685	[illegible]	[illegible]	[illegible]	5746	5761
76	5776	5791	5806	5822	5837	[illegible]	[illegible]	[illegible]	5898	5914
77	5929	5944	5960	5975	5991	6006	[illegible]	6037	6053	6068
78	6084	6100	6115	6131	6147	6162	[illegible]	6194	6209	6225
79	6241	6257	6273	6288	6304	6320	[illegible]	6352	6368	6384
80	6400	6416	6432	6448	6464	6480	[illegible]	6512	6529	6545
81	6561	6577	6593	6610	6626	6642	[illegible]	6675	6691	6708
82	6724	6740	6757	6773	6790	6806	[illegible]	6839	6856	6872
83	6889	6906	6922	6939	6956	6972	[illegible]	7006	7022	7039
84	7056	7073	7090	7106	7123	7140	7157	7174	7191	7208
85	7225	7242	7259	7276	7293	7310	7327	7344	7362	7379
86	7396	7413	7430	7448	7465	7482	[illegible]	7517	7534	7552
87	7569	7586	7604	7621	7639	7656	7674	7691	7709	7726
88	7744	7762	7779	7797	7815	7832	[illegible]	7868	7885	7903
89	7921	7939	7957	7974	7992	8010	[illegible]	8046	8064	8082
90	8100	8118	8136	8154	8172	8190	[illegible]	8226	8245	8263
91	8281	8299	8317	8336	8354	8372	8391	8409	8427	8446
92	8464	8482	8501	8519	8538	8556	[illegible]	8593	8612	8630
93	8649	8668	8686	8705	8724	8742	8761	8780	8798	8817
94	8836	8855	8874	8892	8911	8930	[illegible]	8968	8987	9006
95	9025	9044	9063	9082	9101	9120	[illegible]	9158	9178	9197
96	9216	9235	9254	9274	9293	9312	[illegible]	9351	9370	9390
97	9409	9428	9448	9467	9487	9506	[illegible]	9545	9565	9584
98	9604	9624	9643	9663	9683	9702	[illegible]	9742	9761	9781
99	9801	9821	9841	9860	9880	9900	[illegible]	9940	9960	9980

	0	1	2	3	4	5	6	7	8	9
10	1000 3162	1005 3178	1010 3194	1015 3209	1020 3225	1025 3240	1030 3256	1034 3271	1039 3286	1044 3302
11	1049 3317	1054 3332	1058 3347	1063 3362	1068 3376	1072 3391	1077 3406	1082 3421	1086 3435	1091 3450
12	1095 3464	1100 3479	1105 3493	1109 3507	1114 3521	1118 3536	1122 3550	1127 3564	1131 3578	1136 3592
13	1140 3606	1145 3619	1149 3633	1153 3647	1158 3661	1162 3674	1166 3688	1170 3701	1175 3715	1179 3728
14	1183 3742	1187 3755	1192 3768	1196 3782	1200 3795	1204 3808	1208 3821	1212 3834	1217 3847	1221 3860
15	1225 3873	1229 3886	1233 3899	1237 3912	1241 3924	1245 3937	1249 3950	1253 3962	1257 3975	1261 3987
16	1265 4000	1269 4012	1273 4025	1277 4037	1281 4050	1285 4062	1288 4074	1292 4087	1296 4099	1300 4111
17	1304 4123	1308 4135	1311 4147	1315 4159	1319 4171	1323 4183	1327 4195	1330 4207	1334 4219	1338 4231
18	1342 4243	1345 4254	1349 4266	1353 4278	1356 4290	1360 4301	1364 4313	1367 4324	1371 4336	1375 4347
19	1378 4359	1382 4370	1386 4382	1389 4393	1393 4405	1396 4416	1400 4427	1404 4438	1407 4450	1411 4461
20	1414 4472	1418 4483	1421 4494	1425 4506	1428 4517	1432 4528	1435 4539	1439 4550	1442 4561	1446 4572
21	1449 4583	1453 4593	1456 4604	1459 4615	1463 4626	1466 4637	1470 4648	1473 4658	1476 4669	1480 4680
22	1483 4690	1487 4701	1490 4712	1493 4722	1497 4733	1500 4743	1503 4754	1507 4764	1510 4775	1513 4785
23	1517 4796	1520 4806	1523 4817	1526 4827	1530 4837	1533 4848	1536 4858	1539 4868	1543 4879	1546 4889
24	1549 4899	1552 4909	1556 4919	1559 4930	1562 4940	1565 4950	1568 4960	1572 4970	1575 4980	1578 4990

The first significant figure and the position of the decimal point must be

	0	1	2	3	4	5	6	7	8	9
25	1581 5000	1584 5010	1587 5020	1591 5030	1594 5040	1597 5050	1600 5060	1603 5070	1606 5079	1609 5089
26	1612 5099	1616 5109	1619 5119	1622 5128	1625 5138	1628 5148	1631 5158	1634 5167	1637 5177	1640 5187
27	1643 5196	1646 5206	1649 5215	1652 5225	1655 5235	1658 5244	1661 5254	1664 5263	1667 5273	1670 5282
28	1673 5292	1676 5301	1679 5310	1682 5320	1685 5329	1688 5339	1691 5348	1694 5357	1697 5367	1700 5376
29	1703 5385	1706 5394	1709 5404	1712 5413	1715 5422	1718 5431	1720 5441	1723 5450	1726 5459	1729 5468
30	1732 5477	1735 5486	1738 5495	1741 5505	1744 5514	1746 5523	1749 5532	1752 5541	1755 5550	1758 5559
31	1761 5568	1764 5577	1766 5586	1769 5595	1772 5604	1775 5612	1778 5621	1780 5630	1783 5639	1786 5648
32	1789 5657	1792 5666	1794 5675	1797 5683	1800 5692	1803 5701	1806 5710	1808 5718	1811 5727	1814 5736
33	1817 5745	1819 5753	1822 5762	1825 5771	1828 5779	1830 5788	1833 5797	1836 5805	1838 5814	1841 5822
34	1844 5831	1847 5840	1849 5848	1852 5857	1855 5865	1857 5874	1860 5882	1863 5891	1865 5899	1868 5908
35	1871 5916	1873 5925	1876 5933	1879 5941	1881 5950	1884 5958	1887 5967	1889 5975	1892 5983	1895 5992
36	1897 6000	1900 6008	1903 6017	1905 6025	1908 6033	1910 6042	1913 6050	1916 6058	1918 6066	1921 6075
37	1924 6083	1926 6091	1929 6099	1931 6107	1934 6116	1936 6124	1939 6132	1942 6140	1944 6148	1947 6156
38	1949 6164	1952 6173	1954 6181	1957 6189	1960 6197	1962 6205	1965 6213	1967 6221	1970 6229	1972 6237
39	1975 6245	1977 6253	1980 6261	1982 6269	1985 6277	1987 6285	1990 6293	1992 6301	1995 6309	1997 6317
	2000 6325	2002 6332	2005 6340	2007 6348	2010 6356	2012 6364	2015 6372	2017 6380	2020 6387	2022 6395

	0	1	2	3	4	5	6	7	8	9
41	2025 6403	2027 6411	2030 6419	2032 6427	2035 6434	2037 6442	2040 6450	2042 6458	2045 6465	2047 6473
42	2049 6481	2052 6488	2054 6496	2057 6504	2059 6512	2062 6519	2064 6527	2066 6535	2069 6542	2071 6550
43	2074 6557	2076 6565	2078 6573	2081 6580	2083 6588	2086 6595	2088 6603	2090 6611	2093 6618	2095 6626
44	2098 6633	2100 6641	2102 6648	2105 6656	2107 6663	2110 6671	2112 6678	2114 6686	2117 6693	2119 6701
45	2121 6708	2124 6716	2126 6723	2128 6731	2131 6738	2133 6745	2135 6753	2138 6760	2140 6768	2142 6775
46	2145 6782	2147 6790	2149 6797	2152 6804	2154 6812	2156 6819	2159 6826	2161 6834	2163 6841	2166 6848
47	2168 6856	2170 6863	2173 6870	2175 6877	2177 6885	2179 6892	2182 6899	2184 6907	2186 6914	2189 6921
48	2191 6928	2193 6935	2195 6943	2198 6950	2200 6957	2202 6964	2205 6971	2207 6979	2209 6986	2211 6993
49	2214 7000	2216 7007	2218 7014	2220 7021	2223 7029	2225 7036	2227 7043	2229 7050	2232 7057	2234 7064
50	2236 7071	2238 7078	2241 7085	2243 7092	2245 7099	2247 7106	2249 7113	2252 7120	2254 7127	2256 7134
51	2258 7141	2261 7148	2263 7155	2265 7162	2267 7169	2269 7176	2272 7183	2274 7190	2276 7197	2278 7204
52	2280 7211	2283 7218	2285 7225	2287 7232	2289 7239	2291 7246	2293 7253	2296 7259	2298 7266	2300 7273
53	2302 7280	2304 7287	2307 7294	2309 7301	2311 7308	2313 7314	2315 7321	2317 7328	2319 7335	2322 7342
54	2324 7348	2326 7355	2328 7362	2330 7369	2332 7376	2335 7382	2337 7389	2339 7396	2341 7403	2343 7409
55	2345 7416	2347 7423	2349 7430	2352 7436	2354 7443	2356 7450	2358 7457	2360 7463	2362 7470	2364 7477
56	2366 7483	2369 7490	2371 7497	2373 7503	2375 7510	2377 7517	2379 7523	2381 7530	2383 7537	2385 7543

SQUARE ROOTS

	0	1	2	3	4	5	6	7	8	9
57	2387 7550	2390 7556	2392 7563	2394 7570	2396 7576	2398 7583	2400 7589	2402 7596	2404 7603	2406 7609
58	2408 7616	2410 7622	2412 7629	2415 7635	2417 7642	2419 7649	2421 7655	2423 7662	2425 7668	2427 7675
59	2429 7681	2431 7688	2433 7694	2435 7701	2437 7707	2439 7714	2441 7720	2443 7727	2445 7733	2447 7740
60	2449 7746	2452 7752	2454 7759	2456 7765	2458 7772	2460 7778	2462 7785	2464 7781	2466 7787	2468 7804
61	2470 7810	2472 7817	2474 7823	2476 7829	2478 7838	2480 7842	2482 7849	2484 7855	2486 7861	2488 7863
62	2490 7874	2492 7880	2494 7887	2496 7893	2498 7899	2500 7906	2502 7912	2504 7918	2506 7925	2508 7931
63	2510 7837	2512 7944	2514 7950	2518 7956	2518 7962	2520 7869	2522 7975	2524 7981	2526 7987	2528 7994
64	2530 8000	2532 8006	2534 8012	2536 8019	2538 8025	2540 8031	2542 8037	2544 8044	2546 8050	2548 8056
65	2550 8062	2551 8068	2553 8075	2555 8081	2557 8087	2559 8093	2561 8099	2563 8106	2565 8112	2567 8118
66	2569 8124	2571 8130	2573 8136	2575 8142	2577 8149	2579 8155	2581 8181	2583 8167	2585 8173	2587 8179
67	2588 8185	2590 8191	2592 8198	2594 8204	2596 8210	2598 8216	2600 8222	2602 8228	2604 8234	2606 8240
68	2608 6246	2610 8252	2612 8258	2613 8264	2615 8270	2617 8276	2619 8283	2621 8289	2823 8295	2625 8301
69	2627 8307	2629 8313	2631 8319	2632 8325	2634 8331	2636 8337	2638 8343	8640 8349	2642 8355	2644 8361
70	2646 8367	2648 8373	2650 8379	2651 8385	2653 8390	2655 8396	2657 8402	2659 8408	2661 8414	2663 8420
71	2665 8426	2666 8432	2668 8438	2670 8444	2672 8450	2674 8456	2676 8462	2678 8468	2680 8473	2681 8479
[illegible]		2685 8491	2687 8497	2689 8503	2691 8509	2693 8515	2694 8521	2696 8526	2698 8532	2700 8538

	0	1	2	3	4	5	6	7	8	9
73	2702 8544	2704 8550	2706 8556	2707 8562	2709 8567	2711 8573	2713 8579	2715 8585	2717 8591	2718 8597
74	2720 8602	2722 8608	2724 8614	2726 8620	2728 8626	2729 8631	2731 8637	2733 8643	2735 8649	2737 8654
75	2739 8660	2740 8666	2742 8672	2744 8678	2746 8683	2748 8689	2750 8695	2751 8701	2753 8706	2755 8712
76	2757 8718	2759 8724	2760 8729	2762 8735	2764 8741	2766 8746	2768 8752	2769 8758	2771 8764	2773 8769
77	2775 8775	2777 8781	2778 8786	2780 8792	2782 8798	2784 8803	2786 8809	2787 8815	2789 8820	2791 8826
78	2793 8832	2795 8837	2796 8843	2798 8849	2800 8854	2802 8860	2804 8866	2805 8871	2807 8877	2809 8883
79	2811 8888	2812 8894	2814 8899	2816 8905	2818 8911	2820 8916	2821 8922	2823 8927	2825 8933	2827 8939
80	2828 8944	2830 8950	2832 8955	2834 8961	2835 8967	2837 8972	2839 8978	2841 8983	2843 8989	2844 8994
81	2846 9000	2848 9006	2850 9011	2851 9017	2853 9022	2855 9028	2857 9033	2858 9039	2860 9044	2862 9050
82	2864 9055	2865 9061	2867 9066	2869 9072	2871 9077	2872 9083	2874 9088	2876 9094	2877 9099	2879 9105
83	2881 9110	2883 9116	2884 9121	2886 9127	2888 9132	2890 9138	2891 9143	2893 9149	2895 9154	2897 9160
84	2898 9165	2900 9171	2902 9176	2903 9182	2905 9187	2907 9192	2909 9198	2910 9203	2912 9209	2914 9214
85	2915 9220	2917 9225	2919 9230	2921 9236	2922 9241	2924 9247	2926 9252	2927 9257	2929 9263	2931 9268
86	2933 9274	2934 9279	2936 9284	2938 9290	2939 9295	2941 9301	2943 9306	2944 9311	2946 9317	2948 9322
87	2950 9327	2951 9333	2953 9338	2955 9343	2956 9349	2958 9354	2960 9359	2961 9365	2963 9370	2965 9375
88	2966 9381	2968 9386	2970 9391	2972 9397	2973 9402	2975 9407	2977 9413	2978 9418	2980 9423	2982 9429

	0	1	2	3	4	5	6	7	8	9
89	2983 9434	2985 9439	2987 9445	2988 9450	2990 9455	2992 9460	2993 9466	2995 9471	2997 9476	2998 9482
90	3000 9487	3002 9492	3003 9497	3005 9503	3007 9508	3008 9513	3010 9518	3012 9524	3013 9529	3015 9534
91	3017 9539	3018 9545	3020 9550	3022 9555	3023 9560	3025 9566	3027 9571	3028 9576	3030 9581	3032 9586
92	3033 9592	3035 9597	3036 9602	3038 9607	3040 9612	3041 9618	3043 9623	3045 9628	3046 9633	3048 9638
93	3050 9644	3051 9649	3053 9654	3055 9659	3056 9664	3058 9670	3059 9675	3061 9680	3063 9685	3064 9690
94	3066 9695	3068 9701	3069 9706	3071 9711	3072 9716	3074 9721	3076 9726	3077 9731	3079 9737	3081 9742
95	3082 9747	3084 9752	3085 9757	3087 9762	3089 9767	3090 9772	3092 9778	3094 9783	3095 9788	3097 9793
96	3098 9798	3100 9803	3102 9808	3103 9813	3105 9818	3106 9823	3108 9829	3110 9834	3111 9839	3113 9844
97	3114 9849	3116 9854	3118 9859	3119 9864	3121 9869	3122 9874	3124 9879	3126 9884	3127 9889	3129 9894
98	3130 9899	3132 9905	3134 9910	3135 9915	3137 9920	3138 9925	3140 9930	3142 9935	3143 9940	3145 9945
99	3146 9950	3148 9955	3150 9960	3151 9965	3153 9970	3154 9975	3156 9980	3158 9985	3159 9990	3161 9995

	0	1	2	3	4	5	6	7	8	9
1·0	1·0000	·9901	·9804	·9709	·9615	9524	·9434	·9346	·9259	·9174
1·1	·9091	·9009	·8929	·8850	·8772	8696	·8621	·8547	·8475	·8403
1·2	·8333	·8264	·8197	·8130	·8065	·8000	·7937	7874	·7813	·7752
1·3	·7692	·7634	·7576	·7519	·7463	·7407	·7353	·7299	·7246	·7194
1·4	·7143	·7092	·7042	·6993	·6944	·6897	·6849	·6803	·6757	·6711
1·5	·6667	·6623	·6579	·6536	·6494	·6452	·6410	·6369	·6329	6289
1·6	·6250	·6211	·6173	·6135	·6098	·6061	·6024	·5988	·5952	·5917
1·7	·5882	·5848	·5814	·5780	·5747	·5714	·5682	·5650	5618	·5587
1·8	·5556	·5525	·5495	·5464	5435	·5405	·5376	·5348	·5319	·5291
1·9	·5263	·5236	·5208	·5181	·5155	·5128	·5102	·5076	·5051	·5025
2·0	·5000	·4975	·4950	·4926	·4902	·4878	4854	4831	·4808	·4785
2·1	4762	·4739	·4717	·4695	·4673	·4651	·4630	·4608	·4527	·4566
2·2	·4545	·4525	·4505	·4484	·4464	·4444	·4425	·4405	·4386	·4367
2·3	·4348	·4329	·4310	·4292	·4274	·4255	·4237	·4219	·4202	·4184
2·4	·4167	·4149	·4132	·4115	·4098	·4082	·4065	·4049	·4032	·4016
2·5	·4000	·3984	·3968	·3953	·3937	·3922	·3906	·3891	3876	·3861
2·6	3846	·3831	3817	·3802	·3788	·3774	·3759	·3745	·3731	·3717
2·7	·3704	·3690	·3676	·3663	·3650	·3636	·3623	·3610	3597	·3584
2·8	·3571	·3559	·3546	·3534	·3521	·3509	·2497	·3484	3472	·3460
2·9	·3448	·3436	·3425	·3413	·3401	·3390	·3378	·3367	·3356	·3344
3·0	·3333	·3322	·3311	·3300	·3289	·3279	·3268	·3257	·3247	·3236
3·1	·3226	·3215	·3205	·3195	·3185	·3175	·3165	·3155	·3145	3135
3·2	·3125	·3115	·3106	·3096	·3086	·3077	·3067	3058	·3049	·3040
3·3	·3030	·3021	·3012	·3003	·2994	·2985	·2976	·2967	·2959	·2959
3·4	·2941	·2933	·2924	·2915	·2907	2899	·2890	·2882	·2874	·2865
3·5	·2857	·2849	·2841	·2833	·2825	·2817	·2809	·2801	·2793	·2786
3·6	·2778	·2770	·2762	·2755	·2747	·2740	·2732	·2725	·2717	·2710
3·7	·2703	·2695	·2688	·2681	·2674	·2667	·2660	·2653	·2646	·2639
3·8	·2632	·2625	·2618	·2611	·2604	·2597	·2591	·2584	·2577	2571
3·9	·2564	·2558	·2551	·2545	·2538	·2532	·2525	·2519	·2513	·2506
4·0	·2500	·2494	·2488	·2481	·2475	·2469	·2463	2457	·2451	·2445
4·1	·2439	·2433	·2427	·2421	·2415	·2410	·2404	2398	2392	2387

	0	1	2	3	4	5	6	7	8	9
4·2	2381	2375	2370	2364	2358	2353	2347	2342	2336	2331
4·3	2326	2320	2315	2309	2304	2299	2294	2288	2283	2278
4·4	2273	2268	2262	2257	2252	2247	2242	2237	2292	2227
4·5	2222	2217	2212	2208	2203	2198	2193	2188	2183	2179
4·6	2174	2169	2165	2160	2155	2151	2146	2141	2137	2132
4·7	2128	2123	2119	2114	2110	2105	2101	2096	2092	2088
4·8	2083	2079	2075	2070	2066	2062	[illegible]	2053	[illegible]	[illegible]
4·9	2041	2037	2033	2028	2024	[illegible]	[illegible]	2012	[illegible]	2004
5·0	2000	1996	1992	1988	1984	1980	[illegible]	1972	[illegible]	[illegible]
5·1	1961	1957	1953	1949	1946	1942	[illegible]	1931	1931	1927
5·2	1923	1919	1916	1912	1908	1905	1901	1899	1894	[illegible]
5·3	1887	1883	1880	1876	1873	1869	[illegible]	1862	[illegible]	[illegible]
5·4	1852	1848	1845	1842	1838	1835	[illegible]	1828	[illegible]	1821
5·5	1818	1815	1812	1808	1805	1802	[illegible]	1795	[illegible]	[illegible]
5·6	1786	1783	1779	1776	1773	1770	1767	1764	1761	1757
5·7	1754	1751	1748	1745	1742	1739	[illegible]	1733	[illegible]	1727
5·8	1724	1721	1718	1715	1712	1709	[illegible]	1704	[illegible]	[illegible]
5·9	1695	1692	1689	1686	1684	1681	[illegible]	1675	[illegible]	[illegible]
6·0	1667	1664	1661	1658	1656	1653	[illegible]	1647	[illegible]	[illegible]
6·1	1639	1637	1634	1631	1629	1626	[illegible]	1621	[illegible]	[illegible]
6·2	1613	1610	1608	1605	1603	1600	1597	1595	1592	[illegible]
6·3	1587	1585	1582	1580	1577	1575	1572	1570	1567	[illegible]
6·4	1563	1560	1558	1555	1553	1550	1548	1546	1543	1541
6·5	1538	1536	1534	1531	1529	1527	[illegible]	1522	[illegible]	1517
6·6	1515	1513	1511	1508	1506	1504	[illegible]	1499	1497	[illegible]
6·7	1493	1490	1488	1486	1484	1481	[illegible]	1477	[illegible]	[illegible]
6·8	1471	1468	1466	1464	1462	1460	[illegible]	1456	[illegible]	1451
6·9	1449	1447	1445	1443	1441	1439	[illegible]	1435	[illegible]	1431
7·0	1429	1427	1425	1422	1420	1418	1416	1414	1412	1410
7·1	1408	1406	1404	1403	1401	1399	[illegible]	1395	1393	1391
7·2	1389	1387	1385	1383	1381	1379	1377	1376	1374	1372
7·3	1370	1368	1366	1364	1362	1361	1359	1357	1355	1353
7·4	1351	1350	1348	1346	1344	1342	1340	1339	1337	1335

	0	1	2	3	4	5	6	7	8	9
7·5	·1333	·1332	·1330	·1328	·1326	·1325	·1323	·1321	·1319	·1318
7·6	·1316	·1314	·1312	·1311	·1309	·1307	·1305	·1304	·1302	·1300
7·7	·1299	·1297	·1295	·1294	·1292	·1290	·1289	·1287	·1285	·1284
7·8	·1282	·1280	·1279	·1277	·1276	·1274	·1272	·1271	·1269	·1267
7·9	·1266	·1264	·1263	·1261	·1259	·1258	1256	·1255	·1253	·1252
8·0	·1250	·1248	·1247	·1245	·1244	·1242	·1241	·1239	·1238	·1236
8·1	·1235	·1233	·1232	·1230	·1229	·1227	·1225	·1224	·1222	·1221
8·2	·1220	·1218	·1217	·1215	·1214	·1212	·1211	·1209	·1208	·1206
8·3	·1205	·1203	·1202	·1200	·1199	·1198	·1196	·1195	·1193	·1192
8·4	·1190	·1189	·1188	·1186	·1185	·1183	1182	·1181	·1179	·1178
8·5	·1176	·1175	·1174	·1172	·1171	·1170	·1168	·1167	·1166	·1164
8·6	·1163	·1161	·1160	·1159	·1157	·1156	·1155	·1153	·1152	·1151
8·7	·1149	·1148	·1147	·1145	·1144	·1143	·1142	·1140	·1139	·1138
8·8	·1136	·1135	·1134	·1133	·1131	·1130	·1129	·1127	·1126	·1125
8·9	·1124	·1122	·1121	·1120	·1119	·1117	·1116	·1115	·1114	·1112
9·0	·1111	·1110	·1109	·1107	·1106	·1105	·1104	·1103	·1101	·1100
9·1	·1099	·1098	·1096	·1095	·1094	·1093	·1092	·1091	·1089	·1088
9·2	·1087	·1086	·1085	·1083	·1082	·1081	1080	·1079	·1078	·1076
9·3	·1075	·1074	·1073	·1072	·1071	·1070	·1068	·1067	·1066	·1165
9·4	·1064	·1063	·1062	·1060	·1059	·1058	·1057	·1056	·1055	·1054
9·5	·1053	·1052	·1050	·1049	·1048	·1047	·1046	·1045	·1044	·1043
9·6	·1042	·1041	·1040	·1038	·1037	·1036	·1035	·1034	·1033	·1032
9·7	·1031	·1030	·1029	·1028	·1027	·1026	·1025	·1024	·1022	·1021
9·8	·1020	·1019	·1018	·1017	·1016	·1015	·1014	·1013	·1012	·1011
9·9	·1010	·1009	·1008	·1007	·1006	·1005	·1004	·1003	·1002	·1001

INDEX